

Vishwesh V. Kulkarni · Guy-Bart Stan  
Karthik Raman *Editors*

# A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations

 Springer

# A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations

Vishwesh V. Kulkarni  
Guy-Bart Stan · Karthik Raman  
Editors

# A Systems Theoretic Approach to Systems and Synthetic Biology I: Models and System Characterizations

 Springer

*Editors*

Vishwesh V. Kulkarni  
Electrical and Computer Engineering  
University of Minnesota  
Minneapolis, MN  
USA

Guy-Bart Stan  
Department of Bioengineering  
Imperial College  
London  
UK

Karthik Raman  
Department of Biotechnology  
Bhupat and Jyoti Mehta School  
of Biosciences  
Indian Institute of Technology Madras  
Chennai  
India

ISBN 978-94-017-9040-6      ISBN 978-94-017-9041-3 (eBook)

DOI 10.1007/978-94-017-9041-3

Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2014942544

© Springer Science+Business Media Dordrecht 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))



श्रेयो हि ज्ञानमभ्यासाज्ज्ञानाद्भ्यानं विशिष्यते ।  
ध्यानात् कर्मफलत्यागस्त्यागाच्छान्तिरनन्तरम् ॥ १२-१२ ॥  
– श्रीमद्भगवद्गीता

*Understanding is superior to mere practice  
Union with the subject matter supersedes that  
Dispassion towards all results is better still  
And manifests peace immediately*

Bhagavat Gita (12:12)

*To my father Vasant, brother Vinay, sister  
Ketki, and Prof. Peter Falb  
To my mother in memory*

Vishwesh V. Kulkarni

*To my parents, Florentina and Stephan  
To my wife Cristina and my daughter  
Eva-Victoria*

Guy-Bart Stan

*To my parents and teachers  
In memory of Sunder Mama  
and Prof. E. V. Krishnamurthy*

Karthik Raman

# Foreword

There is no design template more versatile than DNA. Nor are any designs more consequential than those whose blueprints DNA encodes. This exquisite substance has been shaped over billions of years by the creative combination of mutation and selection. Yet in the very long history of this template, it is only during our times that complex living organisms are beginning to understand and manipulate the very template whose sequences define them. But how should we go about this understanding? And how can we use this understanding to more effectively and responsibly alter the DNA template?

The complexity and diversity of living organisms are daunting. *Systems biology* aims at reverse engineering biological complexity for the purpose of understanding their design principles. By measuring and characterizing interactions of key biological molecules in response to stimuli and perturbations, systems biology aims to construct models that capture the complexity of endogenous biological networks. Through the systematic understanding of such models, it is hoped that one will achieve a holistic understanding of biological networks and the way they achieve biological function.

At the same time, the versatility of DNA and the dramatic decrease in the cost of DNA synthesis is making it possible to economically design and test new complex genetic circuits. This has given impetus to a new field: *Synthetic biology*. In our quest to understand biological complexity, we have examined endogenous biological subsystems and ascribed functions and design principles to their components. But a true understanding of these biological design principles is demonstrated only when one can build such systems *de novo* and demonstrate their function. When these circuits do not exhibit behavior consistent with our models, further investigations will lead to a deeper understanding of the underlying biology. Synthetic biology, therefore, serves as an important testbed for our understanding of biological principles. But the promise of synthetic biology extends beyond scientific understanding. Whether it be the detection and interference with the course of disease through the introduction of designer circuits, the cost-effective synthesis of new bio-substances, or the development of improved food products, synthetic biology provides a tremendous opportunity to alleviate suffering and improve the quality of our lives.

In both systems and synthetic biology, challenges abound. Quantitative modeling, analysis, and design of biological networks must contend with difficulties arising from the inescapable fact that at its most basic level, biology involves complex dynamic interactions among nonlinear stochastic components, taking place at multiple temporal and spatial timescales. The complexity of network interconnections of such components and the crosstalk between them adds another level of difficulty.

*System theory* has emerged as a field to deal with the challenges and complexities emerging from the interconnection of engineered systems, many of which are shared with biological systems. Notions from system theory such as nonlinearity, stochasticity, feedback, loading, modularity, robustness, identifiability, etc., are needed for a deeper understanding of biological complexity and for a more reliable design of biological circuits. These concepts are now being utilized to help us expand our understanding of endogenous biological circuits and to design novel ones. The articles in this book make significant strides in this direction.

While system theory will undoubtedly aid our understanding and design of biological systems, there is no doubt that the study of biological designs that have evolved over billions of years will also shape the future of system theory. For example, evolution and development are two central themes in biology that have little analogy with engineered man-made systems. Through the study of these and other biological themes, new systems notions and insights will undoubtedly emerge, enriching system theory in the process. One need only look at the history of *feedback*, a predominant concept in system theory, to imagine what is possible. While its human discovery can be traced back a little over one millennium, it is likely that feedback was invented by nature more than three billion years earlier. Since then, it has been wildly successful as a biological design principle, as evidenced by its prevalence at every level of biological organization. One wonders if an early systematic understanding of this concept in its biological context could have sped up the course of our own technological development.

As the physical sciences helped us understand the physical world around us over the last few centuries, so will quantitative biological science help us understand who we are, how we function, and how we can effectively and responsibly synthesize this most consequential of substances, the DNA. I believe that system theory will be central to this understanding.

Zürich, September 2013

Mustafa Khammash

# Preface

Underlying every living cell are billions of molecules interacting in a beautifully concerted network of pathways such as metabolic, signalling, and regulatory pathways. The complexity of such biological systems has intrigued scientists from many disciplines and has given birth to the highly influential field of *systems biology* wherein a wide array of mathematical techniques, such as flux balance analysis, and technology platforms, such as next generation sequencing, is used to understand, elucidate, and predict the functions of complex biological systems. This field traces its roots to the general systems theory of Ludwig von Bertalanffy and effectively started in 1952 with a mathematical model of the neuronal action potential for which Alan Hodgkin and Andrew Huxley received the Nobel Prize in 1963. More recently, the field of *synthetic biology*, i.e., *de novo* engineering of biological systems, has emerged. Here, the phrase ‘biological system’ can assume a vast spectrum of meanings: DNA, protein, genome, cell, cell population, tissue, organ, ecosystem, and so on. Scientists from various fields are focusing on how to render this *de novo* engineering process more predictable, reliable, scalable, affordable, and easy. Systems biology and synthetic biology are essentially two facets of the same entity. As was the case with electronics research in the 1950s, a large part of synthetic biology research, such as the *BioFab* project, has focused on reusable macromolecular “parts” and their standardization so that composability can be guaranteed. Recent breakthroughs in DNA synthesis and sequencing combined with newly acquired means to synthesize plasmids and genomes have enabled major advances in science and engineering and marked the true beginning of the era of synthetic biology. Significant industrial investments are already underway. For example, in 2009, Exxon Mobil set up a collaboration worth \$600 million with Synthetic Genomics to develop next generation biofuels.

Recent advances in systems and synthetic biology clearly demonstrate the benefits of a rigorous and systematic approach rooted in the principles of systems and control theory—not only does it lead to exciting insights and discoveries but it also reduces the inordinately lengthy trial-and-error process of wet-lab experimentation, thereby facilitating significant savings in human and financial resources. So far, state-of-the-art systems-and-control-theory-inspired results in systems and synthetic biology have been scattered across various books and journals from various disciplines. Hence, we felt the need for an edited book that provides a

panoramic view and illustrates the potential of such systematic and rigorous mathematical methods in systems and synthetic biology.

Systems and control theory is a branch of engineering and applied sciences that rigorously deals with the complexities and uncertainties of interconnected systems with the objective of characterizing fundamental systemic properties such as stability, robustness, communication capacity, and other performance metrics. Systems and control theory also strives to offer concepts and methods that facilitate the design of systems with rigorous guarantees on these fundamental properties. For more than 100 years, the insights and techniques provided by systems and control theory have enabled outstanding technological contributions in diverse fields such as aerospace, telecommunication, storage, automotive, power systems, and others. Notable examples include Lyapunov's theorems, Bellman's theory of dynamic programming, Kalman's filter,  $H^\infty$  control theory, Nyquist-Shannon sampling theorem, Pontryagin's minimum principle, and Bode's sensitivity integral. Can systems and control theory have, or evolve to have, a similar impact in biology? The chapters in this book demonstrate that, indeed, systems and control theoretic concepts and techniques can be useful in our quest to understand how biological systems function and/or how they can be (re-)designed from the bottom up to yield new biological systems that have rigorously characterized robustness and performance properties.

Several barriers must be overcome to contribute significantly in this exciting journey. One of these is the language barrier, e.g., what a systems theorist means by the word *sensitivity* is different from what a biologist means by it. Another one is the knowledge barrier as, traditionally, systems and control theorists and biologists are not well versed with each other's knowledge base (although that scenario is now fast changing for the better with the introduction of bioengineering courses in systems and control theory at the undergraduate and graduate levels). A third barrier is due to the sheer volume of *big data*: the European Bioinformatics Institute in Hixton, UK, which is one of the world's largest biological data repositories, currently stores 20 petabytes of data and backups about genes, proteins and small molecules, and this number is more than doubling every year. Finally, a fourth barrier comes from the effort required to produce timely contributions based on currently available models. As an example of this last barrier, the systems and control theory community could have played a greater role than it did in two of the most significant technological advances of the last 50 years: VLSI and Internet. In retrospect, besides the fact that the systems and control theorists caught on to the Internet too late, by which time infrastructures based on TCP/IP were already in place, the main difficulty posed by the Internet for the systems and control theory community was a lack of *good* models of the underlying networked system. This lack-of-good-models barrier is even more daunting in biology since some of the currently available *big data* are not guaranteed to be reproducible. As Prof. M. Vidyasagar illustrates and observes in the September 2012 issue of IEEE Lifesciences, one of the major challenges to the application of systems and control theory concepts in biology comes from "the fact that many biological experiments are not fully repeatable, and thus the resulting data sets are not readily amenable to

the application of methods that people like us [i.e., systems and control theorists] take for granted.”

The chapters in this book serve to propose ways to overcome such barriers and to illustrate that biologists as well as systems and control theorists can make deep and timely contributions in life sciences by collaborating with each other to solve important questions such as how to devise experiments to obtain models of biological systems, how to obtain predictive models using information extracted from experimental data, how to choose components for (re-)engineering biological networks, how to adequately interconnect biological systems, and so on. Furthermore, and as Prof. Mustafa Khammash observes in his foreword, this research will fundamentally enrich systems and control theory as well by forcing it to investigate currently open questions that are specific to living biological systems, e.g., Why do biological systems naturally evolve the way they do? Can the evolvability of biological systems be consciously exploited for (re-)design and optimization purposes?

This book is intended for (1) systems and control theorists interested in molecular and cellular biology, and (2) biologists interested in rigorous modeling, analysis, and control of biological systems. We believe that research at the intersection of these disciplines will foster exciting discoveries and will stimulate mutually beneficial developments in systems & control theory and systems & synthetic biology.

The book consists of 12 chapters contributed by leading researchers from the fields of systems and control theory, systems biology, synthetic biology, and computer science. Chapters 1–6 focus on general mathematical concepts, methods, and tools that are currently used to answer important questions in biology. Chapters 7–12 describe various biological network modeling approaches used to untangle biological complexity and reverse-engineer biological networks from data.

- **Part I—Mathematical Analysis:** Chapters 1–6 present core mathematical concepts and methods that can be used and further adapted for solving specific problems in biology. As an example, consider the law of mass action. It has been widely used in chemistry since Guldberg and Waage formulated it in 1864. But does it have a deeper significance that is applicable outside chemistry? Likewise, reaction-diffusion systems feature in all pattern formation problems which, in turn, are significant in neuronal networks and disease phenotypes. Under which conditions is spatial uniformity guaranteed? The chapters in this part provide rigorous mathematical foundations that can be used to resolve such questions. A brief summary of each chapter is as follows.

- Chapter 1: The law of mass action is used in (bio-)chemistry to characterize and predict the behavior of interacting (bio-)chemical species. Guldberg and Waage formulated it in 1864 and it has since been built upon and widely used in (bio-)chemistry and cellular biology. To make it available for consideration by researchers in areas other than chemistry, Adleman et al. present it in a new form, viz., in the context of event systems, after solidifying its mathematical foundations.

- Chapter 2: Molecular systems often have a mathematical representation with uncertainties embedded in it. These uncertainties make predictions of the system’s behavior harder. Nonetheless, it is still possible, in some scenarios, to obtain certain qualitative behavioral results that are fairly parameter independent and, instead, are a property of the system structure. Blanchini and Franco use a parameter-free qualitative modeling framework and show under which conditions behaviors such as oscillations and multi-stability are only structure dependent.
  - Chapter 3: Reaction-diffusion systems are of central importance in all applications that feature pattern formations. Aminzare et al. present conditions that guarantee the spatial uniformity of the solutions of reaction-diffusion partial differential equations. They demonstrate that these conditions can be verified using linear matrix inequalities and outline the applicability of these results in analyzing biological oscillations and enzymatic signalling pathways.
  - Chapter 4: Biologists often rely on linearized models to examine stability and on phase-plane analysis to understand the effect of parameter variations. Although useful, phase-plane analysis cannot be used to address simultaneous variations in more than two parameters. Kulkarni et al. show how multiplier theory can be used to overcome these limitations and illustrate its use via a case study of the celebrated Elowitz–Leibler oscillator.
  - Chapter 5: Modularity possibly emerged at the cellular level through natural selection and evolution. But do modules make sense in the context of metabolic networks? Goelzer and Fromion present a framework that allows a modular decomposition of steady-state metabolic networks, and show how this framework can also be used for a qualitative predictive modeling based on omics datasets.
  - Chapter 6: Biological network modeling often encounters the problem of how to deal with hidden state dynamics. Santiello et al. address the problem of predicting hidden state transitions from temporal sequential datasets (for example, EEG, EMG, MER) by developing a Bayesian detection paradigm that combines optimal control and Markov processes.
- **Part II—Biological Network Modeling:** Chapters 7–12 focus on certain techniques that can be used to obtain predictive models of biological networks. Here, the limitations of the perturbation methods used to generate the data, the vast amount of available data (which does not necessarily correlate with the amount of *information* they contain), hidden states, measurement noise, and other factors combine to render this broad area of research one of the greatest scientific and technological challenges of today. The chapters in this section summarize some of these challenges and present architectures that constitute an important step in arriving at a definitive solution. Somewhat similar, but less complex system identification problems have been encountered and resolved in systems theory and computer science over the last decades. Can these techniques and the insight they provide be useful in biology? To answer this



question, it is crucial to understand the advantages and limitations of each particular technique. The set of chapters collated in this part aim to highlight the current state of the art for biological network modeling and the advantages and limitations of the presented approaches. A brief summary of each chapter is as follows.

- Chapter 7: In metabolic networks, the metabolite dynamics evolve on much shorter timescales than their catalytic enzymes. Kuntz et al. show how such timescale separation can be exploited using Tikhonov’s theorem for singularly perturbed systems to derive reduced models whose behaviors are guaranteed to remain quantifiably close to those of the non-reduced models. They illustrate this approach by applying it to an example of genetic feedback control for branched metabolic pathways.
- Chapter 8: A central theme in complex network theory, popularized by the study of small-world and scale-free networks at the turn of the last century, is the study of biological networks using various metrics. In this chapter, Roy discusses the utility of various network metrics as well as the need to go beyond fundamental metrics, such as node degree, to better understand how an organism’s phenotype is encoded by its network topology.
- Chapter 9: Even though most of the complex real-world systems exhibit nonlinearities, linear models serve as a useful first order approximation. Carignano et al. present a detailed exposition on how linear system identification techniques can be used to obtain causal relationships between biomolecular entities.
- Chapter 10: Fisher and Piterman discuss how ideas from computer science can be useful for *model checking* in systems biology. They present a methodology to analyze biochemical networks, and specifically a method to test for a faithful reproduction of biological interactions that are known a priori as well as to identify interactions that are not known a priori.
- Chapter 11: Bussetto et al. discuss objective-specific strategies for designing informative experiments in systems biology. Following a formal description of the task of experimental design, they illustrate the use of Bayesian and information-theoretic approaches to design experiments in systems biology.
- Chapter 12: Today, there is a critical need for new methods that rapidly transform high-throughput genomics, transcriptomics, and metabolomics data into predictive network models for metabolic engineering and synthetic biology. In this chapter, Chandrasekaran describes the state of the art of these methods and explains an approach for this purpose called Probabilistic Regulation of Metabolism (PROM).

The burgeoning fields of systems biology and synthetic biology have thrown up a very large number of interesting research problems. As the pre-eminent computer scientist Donald Knuth put it, “biology easily has 500 years of exciting problems to work on.” The chapters in this book address but a small fraction of these interesting challenges. Nevertheless, we believe this book can serve as a

good introduction on some of the currently open problems and on some of the state-of-the-art concepts and techniques available to propose solutions to such problems.

We are very grateful to all authors for their invaluable time and contributions and to Prof. Mustafa Khammash (ETH Zürich) for his stimulating foreword. We are also grateful to our institutions: University of Minnesota (Minneapolis, USA), Imperial College (London, UK), and Indian Institute of Technology Madras (Chennai, India) for their support and for providing a stimulating work environment. Finally, we thank and acknowledge the financial support of our respective funding agencies: the National Science Foundation, the UK Engineering and Physical Sciences Research Council, and the Ministry of Human Resource and Development of the Government of India.

Minneapolis, MN, USA, September 2013  
London, UK  
Chennai, India

Vishwesh V. Kulkarni  
Guy-Bart Stan  
Karthik Raman

# Contents

## Part I Mathematical Analysis

<b>1</b>	<b>On the Mathematics of the Law of Mass Action</b> . . . . .	3
	Leonard Adleman, Manoj Gopalkrishnan, Ming-Deh Huang, Pablo Moisset and Dustin Reishus	
<b>2</b>	<b>Structural Analysis of Biological Networks</b> . . . . .	47
	Franco Blanchini and Elisa Franco	
<b>3</b>	<b>Guaranteeing Spatial Uniformity in Reaction-Diffusion Systems Using Weighted <math>L^2</math> Norm Contractions.</b> . . . . .	73
	Zahra Aminzare, Yusef Shafi, Murat Arcak and Eduardo D. Sontag	
<b>4</b>	<b>Robust Tunable Transcriptional Oscillators Using Dynamic Inversion</b> . . . . .	103
	Vishwesh V. Kulkarni, Aditya A. Paranjape and Soon-Jo Chung	
<b>5</b>	<b>Towards the Modular Decomposition of the Metabolic Network.</b> . . . . .	121
	Anne Goelzer and Vincent Fromion	
<b>6</b>	<b>An Optimal Control Approach to Seizure Detection in Drug-Resistant Epilepsy</b> . . . . .	153
	Sabato Santaniello, Samuel P. Burns, William S. Anderson and Sridevi V. Sarma	

## Part II Biological Network Modelling

<b>7</b>	<b>Model Reduction of Genetic-Metabolic Networks via Time Scale Separation</b> . . . . .	181
	Juan Kuntz, Diego Oyarzún and Guy-Bart Stan	

**8 Networks, Metrics, and Systems Biology . . . . . 211**  
Soumen Roy

**9 Understanding and Predicting Biological Networks Using  
Linear System Identification . . . . . 227**  
Alberto Carignano, Ye Yuan, Neil Dalchau, Alex A. R. Webb  
and Jorge Gonçalves

**10 Model Checking in Biology . . . . . 255**  
Jasmin Fisher and Nir Piterman

**11 Computational Design of Informative Experiments  
in Systems Biology . . . . . 281**  
Alberto Giovanni Busetto, Mikael Sunnåker  
and Joachim M. Buhmann

**12 Predicting Phenotype from Genotype Through Reconstruction  
and Integrative Modeling of Metabolic and Regulatory  
Networks . . . . . 307**  
Sriram Chandrasekaran

**Index . . . . . 327**

**Part I**  
**Mathematical Analysis**

# Chapter 1

## On the Mathematics of the Law of Mass Action

Leonard Adleman, Manoj Gopalkrishnan, Ming-Deh Huang, Pablo Moisset and Dustin Reishus

**Abstract** In 1864, Waage and Guldberg formulated the “law of mass action.” Since that time, chemists, chemical engineers, physicists and mathematicians have amassed a great deal of knowledge on the topic. In our view, sufficient understanding has been acquired to warrant a formal mathematical consolidation. A major goal of this consolidation is to solidify the mathematical foundations of mass action chemistry—to provide precise definitions, elucidate what can now be proved, and indicate what is only conjectured. In addition, we believe that the law of mass action is of intrinsic mathematical interest and should be made available in a form that might transcend its application to chemistry alone. We present the law of mass action in the context of a dynamical theory of sets of binomials over the complex numbers.

**Keywords** Law of mass action · Mass action kinetics · Event systems · Binomials · String theory · Differential equations · Flow-invariant affine subspaces · Natural event systems · Lyapunov function

---

L. Adleman · M.-D. Huang  
University of Southern California, Los Angeles, CA, USA  
e-mail: adleman@usc.edu

M.-D. Huang  
e-mail: huang@usc.edu

M. Gopalkrishnan (✉)  
Tata Institute of Fundamental Research, Mumbai, India  
e-mail: manojg@tifr.res.in

P. Moisset  
Universidad de Chile, Santiago, Chile  
e-mail: pablo.moisset@gmail.com

D. Reishus  
University of Colorado Boulder, Boulder, CO, USA  
e-mail: reishus@colorado.edu

## 1.1 Introduction

The study of mass action kinetics dates back at least to 1864, when Guldberg and Waage [7] formulated the “law of mass action.” Since that time, a great deal of knowledge on the topic has been amassed in the form of empirical facts, physical theories and mathematical theorems by chemists, chemical engineers, physicists and mathematicians. In recent years, Horn and Jackson [10], and Feinberg [5] have made significant mathematical contributions, and these have guided our work.

It is our view that a critical mass of knowledge has been obtained, sufficient to warrant a formal mathematical consolidation. A major goal of this consolidation is to solidify the mathematical foundations of this aspect of chemistry—to provide precise definitions, elucidate what can now be proved, and indicate what is only conjectured. In addition, we believe that the law of mass action is of intrinsic mathematical interest and should be made available in a form that might transcend their application to chemistry alone.

To make the law of mass action available for consideration by researchers in areas other than chemistry, we present mass action kinetics in a new form, which we call event-systems. Our formulation begins with the observation that systems of chemical reactions can be represented by sets of binomials. This gives us an opportunity to extend the law of mass action to arbitrary sets of binomials. Once this extension is made, there is no reason to restrict ourselves to binomials with real coefficients. Hence, we are led to a dynamical theory of sets of binomials over the complex numbers. Possible mathematical applications of this theory include:

1. Binomials are objects of intrinsic mathematical interest [4]. For example, they occur in the study of toric varieties, and hence in string theory. With each set of binomials over the complex numbers, we associate a corresponding system of differential equations. Ideally, this dynamical viewpoint will help advance the theory of binomials, and enhance our understanding of their associated algebraic sets.
2. When we extend the study of the law of mass action to sets of binomials over the complex numbers, we can consider reactions that involve complex rates, complex concentrations, and move through complex time. Extending to the complex numbers gives us direct access to the powerful theorems of complex analysis. Though this clearly transcends conventional chemistry, it may have applications in pure mathematics.

For example, in ongoing work, we seek to exploit an analogy between number theory and chemistry, where atoms are to molecules as primes are to numbers. We associate a distinct species with each natural number. Then each multiplication rule  $m \times n = mn$  is encoded by a reaction where the species corresponding to the number  $m$  reacts with the species corresponding to the number  $n$  to form the species corresponding to the number  $mn$ . With an appropriate choice of specific rates of reactions the resulting event-system has the property that the sum of equilibrium concentrations of all species at complex temperature  $s$  is the value of the

Riemann zeta function at  $s$ . We hope to pursue this approach to study questions related to the distribution of the primes.

3. Systems of linear differential equations are well understood. In contrast, systems of ordinary non-linear differential equations can be notoriously intractable. Differential equations that arise from event-systems lie somewhere in between—more structured than arbitrary non-linear differential equations, but more challenging than linear differential equations. As such, they appear to be an important new class for consideration in the theory of ordinary differential equations.

In addition to their use in mathematics, event-systems provide a vehicle by which ideas in algebraic geometry may be made readily available to the study of mass action kinetics. As such, they may help solidify the foundations of this aspect of chemistry. We expand on this in Sect. 1.7.

Part of our motivation for this research comes from the emerging field of nanotechnology. To quote from [1], “Self-assembly is the ubiquitous process by which objects autonomously assemble into complexes. Nature provides many examples: Atoms react to form molecules. Molecules react to form crystals and supramolecules. Cells sometimes coalesce to form organisms. Even heavenly bodies self-assemble into astronomical systems. It has been suggested that self-assembly will ultimately become an important technology, enabling the fabrication of great quantities of small objects such as computer circuits... Despite its importance, self-assembly is poorly understood.” Hopefully, the theory of event-systems is a step towards understanding this important process.

The chapter is organized as follows:

In Sect. 1.2, we present the basic mathematical notations and definitions for the study of event-systems.

In Sect. 1.3, and all of the sections that follow, we restrict to finite event-systems. Theorem 3 demonstrates that the stoichiometric coefficients give rise to flow-invariant affine subspaces—“conservation classes.”

In Sect. 1.4, and all of the sections that follow, we restrict to “physical event-systems.” Though we have defined event-systems over the complex numbers, in this chapter we focus on consolidating results from the mass action kinetics of reversible chemical reactions. Physical event-systems capture the idea that the specific rates of chemical reactions are always positive real numbers. The main result of this section is Theorem 4, which demonstrates that for physical event-systems, if initially all concentrations are non-negative, then they stay non-negative for all future real times so long as the solution exists. Further, the concentration of every species whose initial concentration is positive, stays positive.

In Sect. 1.5, and all the sections that follow, we restrict to “natural event-systems.” Natural event-systems capture the concept of detailed balance from chemistry. In Theorem 5, we give four equivalent characterizations of natural event-systems; in particular, we show that natural event-systems are precisely those physical event-systems that have no “energy cycles.” In Theorem 7, following Horn and Jackson [10], we show that natural event-systems have associated Lyapunov functions. This theorem is reminiscent of the second law of thermodynamics. The main result



of this section is Theorem 10, which establishes that for natural event-systems, given non-negative initial conditions:

1. Solutions exist for all forward real times.
2. Solutions are uniformly bounded in forward real time.
3. All positive equilibria satisfy detailed balance.
4. Every conservation class containing a positive point also contains exactly one positive equilibrium point.
5. Every positive equilibrium point is asymptotically stable relative to its conservation class.

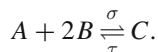
For systems of reversible reactions that satisfy detailed balance, must concentrations approach equilibrium? We believe this to be the case, but are unable to prove it. In 1972, an incorrect proof was offered [10, Lemma 4C]. This proof was retracted in 1974 [9]. To the best of our knowledge, this question in mass action kinetics remains unresolved [14, p. 10]. We pose it formally in Open Problem 1, and consider it the fundamental open question in the field.

In Sect. 1.6, we introduce the notion of “atomic event-systems.” As the name suggests, this is an attempt to capture mathematically the atomic hypothesis that all species are composed of atoms. The main theorem of this section is Theorem 11, which establishes that for natural, atomic event-systems, solutions with positive initial conditions asymptotically approach positive equilibria. Hence, Open Problem 1 is resolved in the affirmative for this restricted class of event-systems.

## 1.2 Basic Definitions and Notation

Before formally defining event-systems, we give a very brief, informal introduction to chemical reactions. All reactions are assumed to take place at constant temperature in a well-stirred vessel of constant volume.

Consider



This chemical equation concerns the reacting species  $A$ ,  $B$  and  $C$ . In the forward direction, one mole of  $A$  combines with two moles of  $B$  to form one mole of  $C$ . The symbol “ $\sigma$ ” represents a real number greater than zero. It denotes, in appropriate units, the rate of the forward reaction when the reaction vessel contains one mole of  $A$  and one mole of  $B$ . It is called the specific rate of the forward reaction. In the reverse direction, one mole of  $C$  decomposes to form one mole of  $A$  and two moles of  $B$ . The symbol “ $\tau$ ” represents the specific rate of the reverse reaction. Chemists typically determine specific rates empirically. Though irreversible reactions (those with  $\sigma = 0$  or  $\tau = 0$ ) have been studied, they will not be considered in this chapter.

Inspired by the law of mass action, we introduce a multiplicative notation for chemical reactions, as an alternative to the chemical equation notation. In our

notation, each chemical reaction is represented by a binomial. Consider the following examples. On the left are chemical equations. On the right are the corresponding binomials.

$$\begin{aligned}
 X_2 \xrightleftharpoons[1/3]{1/2} X_1 &\rightarrow \frac{1}{3}X_2 - \frac{1}{2}X_1 \\
 X_3 \xrightleftharpoons[1/3]{1/2} X_1 + X_2 &\rightarrow \frac{1}{3}X_3 - \frac{1}{2}X_1X_2 \\
 2X_1 + 3X_6 \xrightleftharpoons[\tau]{\sigma} 3X_1 + 2X_2 &\rightarrow \sigma X_1^2X_6^3 - \tau X_1^3X_2^2
 \end{aligned}$$

Our notation leads us to view every set of binomials over an arbitrary field  $\mathbb{F}$  as a formal system of reversible reactions with specific rates in  $\mathbb{F} \setminus \{0\}$ . For our present purposes, we will restrict our attention to binomials over the complex numbers. With this in mind, we now define our notion of event-system.

**Notation 1** Let  $\mathbb{C}_\infty = \bigcup_{n=1}^\infty \mathbb{C}[X_1, X_2, \dots, X_n]$ . A monic monomial of  $\mathbb{C}_\infty$  is a product of the form  $\prod_{i=1}^\infty X_i^{e_i}$  where the  $e_i$  are non-negative integers all but finitely many of which are zero. We will write  $\mathbb{M}_\infty$  to denote the set of all monic monomials of  $\mathbb{C}_\infty$ . More generally, if  $S \subset \{X_1, X_2, \dots\}$ , we let  $\mathbb{C}[S]$  be the ring of polynomials with indeterminants in  $S$  and we let  $\mathbb{M}_S = \mathbb{M}_\infty \cap \mathbb{C}[S]$  (i.e. the monic monomials in  $\mathbb{C}[S]$ ).

If  $n \in \mathbb{Z}_{>0}$ ,  $p \in \mathbb{C}[X_1, X_2, \dots, X_n]$ , and  $\mathbf{a} = \langle a_1, a_2, \dots, a_n \rangle \in \mathbb{C}^n$  then, as is usual, we will let  $p(\mathbf{a})$  denote the value of  $p$  on argument  $\mathbf{a}$ .

Given two monic monomials  $M = \prod_{i=1}^\infty X_i^{e_i}$  and  $N = \prod_{i=1}^\infty X_i^{f_i}$  from  $\mathbb{M}_\infty$ , we will say  $M$  precedes  $N$  (and we will write  $M < N$ ) iff  $M \neq N$  and for the least  $i$  such that  $e_i \neq f_i$ ,  $e_i < f_i$ .

It follows that 1 is a monic monomial of  $\mathbb{C}_\infty$  and that each element of  $\mathbb{C}_\infty$  is a  $\mathbb{C}$ -linear combination of finitely many monic monomials. We will be particularly concerned with the set of binomials  $\mathbb{B}_\infty = \{\sigma M + \tau N \mid \sigma, \tau \in \mathbb{C} \setminus \{0\} \text{ and } M, N \text{ are distinct monic monomials of } \mathbb{C}_\infty\}$ .

**Definition 2** (Event-system) An event-system  $\mathcal{E}$  is a nonempty subset of  $\mathbb{B}_\infty$ .

If  $\mathcal{E}$  is an event-system, its elements will be called “ $\mathcal{E}$ -events” or just “events.” Note that if  $\sigma M + \tau N$  is an event then  $M \neq N$ .

Our map from chemical equations to events is as follows. A chemical equation

$$\begin{aligned}
 \sum_i a_i X_i \xrightleftharpoons[\tau]{\sigma} \sum_j b_j X_j &\text{ goes to:} \\
 1. \sigma \prod_i X_i^{a_i} - \tau \prod_j X_j^{b_j} &\text{ if } \prod_i X_i^{a_i} < \prod_j X_j^{b_j} \\
 \text{or } 2. \tau \prod_j X_j^{b_j} - \sigma \prod_i X_i^{a_i} &\text{ if } \prod_j X_j^{b_j} < \prod_i X_i^{a_i}
 \end{aligned}$$

For example:

$$\begin{aligned}
 X_1 &\stackrel{1/3}{\underset{1/2}{\rightleftharpoons}} X_2 \rightarrow \frac{1}{3}X_2 - \frac{1}{2}X_1 \quad (\text{because } X_2 < X_1) \\
 X_2 &\stackrel{1/2}{\underset{1/3}{\rightleftharpoons}} X_1 \rightarrow \frac{1}{3}X_2 - \frac{1}{2}X_1 \\
 X_1 &\stackrel{-1/3}{\underset{-1/2}{\rightleftharpoons}} X_2 \rightarrow -\frac{1}{3}X_2 + \frac{1}{2}X_1 \\
 X_1 &\stackrel{1/3}{\underset{-1/2}{\rightleftharpoons}} X_2 \rightarrow \frac{1}{3}X_2 + \frac{1}{2}X_1 \\
 X_1 + X_2 &\stackrel{1/3}{\underset{1/2}{\rightleftharpoons}} X_3 \rightarrow \frac{1}{3}X_3 - \frac{1}{2}X_1X_2 \\
 3X_1 + 2X_2 &\stackrel{\sigma}{\underset{\tau}{\rightleftharpoons}} 2X_1 + 3X_6 \rightarrow \tau X_1^2 X_6^3 - \sigma X_1^3 X_2^2
 \end{aligned}$$

Note that our order of monomials is arbitrary. Any linear order would do. The order is necessary to achieve a one-to-one map from chemical reactions to events.

Our definition of event-systems allows for an infinite number of reactions, and an infinite number of reacting species. Indeed, polymerization reactions are commonplace in nature and, in principle, they are capable of creating arbitrarily long polymers (for example, DNA molecules).

The next definition introduces the notion of systems of reactions for which the number of reacting species is finite.

**Definition 3** (*Finite-dimensional event-system*) An event-system  $\mathcal{E}$  is *finite-dimensional* iff there exists an  $n \in \mathbb{Z}_{>0}$  such that  $\mathcal{E} \subset \mathbb{C}[X_1, X_2, \dots, X_n]$ .

**Definition 4** (*Dimension of event-systems*) Let  $\mathcal{E}$  be a finite-dimensional event-system. Then the least  $n$  such that  $\mathcal{E} \subset \mathbb{C}[X_1, X_2, \dots, X_n]$  is the *dimension* of  $\mathcal{E}$ .

**Definition 5** (*Physical event, Physical event-system*) A binomial  $e \in \mathbb{B}_\infty$  is a *physical event* iff there exist  $\sigma, \tau \in \mathbb{R}_{>0}$  and  $M, N \in \mathbb{M}_\infty$  such that  $M < N$  and  $e = \sigma M - \tau N$ . An event-system  $\mathcal{E}$  is *physical* iff each  $e \in \mathcal{E}$  is physical.

Chemical reaction systems typically have positive real forward and backward rates. Physical event-systems generalize this notion.

**Definition 6** Let  $n \in \mathbb{Z}_{>0}$ . Let  $\alpha = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle \in \mathbb{C}^n$ .

1.  $\alpha$  is a *non-negative point* iff for  $i = 1, 2, \dots, n$ ,  $\alpha_i \in \mathbb{R}_{\geq 0}$ .
2.  $\alpha$  is a *positive point* iff for  $i = 1, 2, \dots, n$ ,  $\alpha_i \in \mathbb{R}_{>0}$ .
3.  $\alpha$  is a *z-point* iff there exists an  $i$  such that  $\alpha_i = 0$ .

In chemistry, a system is said to have achieved detailed balance when it is at a point where the net flux of each reaction is zero. Given the corresponding event-system, points of detailed balance corresponds to points where each event evaluates to zero, and vice versa. We call such points “strong equilibrium points.”

**Definition 7** (*Strong equilibrium point*) Let  $\mathcal{E}$  be a finite-dimensional event-system of dimension  $n$ .  $\alpha \in \mathbb{C}^n$  is a *strong  $\mathcal{E}$ -equilibrium point* iff for all  $e \in \mathcal{E}$ ,  $e(\alpha) = 0$ .

In the language of algebraic geometry, when  $\mathcal{E}$  is a finite-dimensional event-system, its corresponding algebraic set is precisely the set of its strong  $\mathcal{E}$ -equilibrium points.

It is widely believed that all “real” chemical reactions achieve detailed balance. We now introduce natural event-systems, a restriction of finite-dimensional, physical event-systems to those that can achieve detailed balance.

**Definition 8** (*Natural event-system*) A finite-dimensional event-system  $\mathcal{E}$  is *natural* iff it is physical and there exists a positive strong  $\mathcal{E}$ -equilibrium point.

Our next goal is to introduce atomic event-systems: finite-dimensional event-systems obeying the atomic hypothesis that all species are composed of atoms. Towards this goal, we will define a graph for each finite-dimensional event-system. The vertices of this graph are the monomials from  $\mathbb{M}_\infty$  and the edges are determined by the events. If a weight  $r$  is assigned to an edge, then  $r$  represents the energy released when a reaction corresponding to that edge takes place. For the purpose of defining atomic event-systems, the reader may ignore the weights; they are included here for use elsewhere in the chapter (Definition 24).

Though graphs corresponding to systems of chemical reactions have been defined elsewhere (e.g. [5, 14, p. 10]), it is important to note that these definitions do not coincide with ours.

**Definition 9** (*Event-graph*) Let  $\mathcal{E}$  be a finite-dimensional event-system. The event-graph  $G_\mathcal{E} = \langle V, E, w \rangle$  is a weighted, directed multigraph such that:

1.  $V = \mathbb{M}_\infty$
2. For all  $M_1, M_2 \in \mathbb{M}_\infty$ , for all  $r \in \mathbb{C}$ ,  $\langle M_1, M_2 \rangle \in E$  and  $r \in w(\langle M_1, M_2 \rangle)$  iff there exist  $e \in \mathcal{E}$  and  $\sigma, \tau \in \mathbb{C}$  and  $M, N, T \in \mathbb{M}_\infty$  such that  $e = \sigma M + \tau N$  and  $M < N$  and either
  - (a)  $M_1 = TM$  and  $M_2 = TN$  and  $r = \ln\left(-\frac{\sigma}{\tau}\right)$  or
  - (b)  $M_1 = TN$  and  $M_2 = TM$  and  $r = -\ln\left(-\frac{\sigma}{\tau}\right)$

Notice that two distinct weights  $r_1$  and  $r_2$  could be assigned to a single edge. For example, let  $\mathcal{E} = \{X_1X_2 - 2X_1^2, X_2 - 5X_1\}$ . Consider the edge in  $G_\mathcal{E}$  from the monomial  $X_1^2$  to the monomial  $X_1X_2$ . Weight  $\ln 2$  is assigned to this edge due to the event  $X_1X_2 - 2X_1^2$ , with  $T = 1$ . Weight  $\ln 5$  is also assigned to this edge due to the event  $X_2 - 5X_1$ , with  $T = X_1$ .

**Definition 10** Let  $\mathcal{E}$  be a finite-dimensional event-system. For all  $M \in \mathbb{M}_\infty$ , the *connected component* of  $M$ , denoted  $C_\mathcal{E}(M)$ , is the set of all  $N \in \mathbb{M}_\infty$  such that there is a path in  $G_\mathcal{E}$  from  $M$  to  $N$ .

It follows from the definition of “path” that every monomial belongs to its connected component.

**Definition 11** (*Atomic event-system*) Let  $\mathcal{E}$  be a finite-dimensional event-system of dimension  $n$ . Let  $S = \{X_1, X_2, \dots, X_n\}$ . Let  $A_{\mathcal{E}} = \{X_i \in S \mid C_{\mathcal{E}}(X_i) = \{X_i\}\}$ .  $\mathcal{E}$  is *atomic* iff for all  $M \in \mathbb{M}_S$ ,  $C(M)$  contains a unique monomial in  $\mathbb{M}_{A_{\mathcal{E}}}$ .

If  $\mathcal{E}$  is atomic then the members of  $A_{\mathcal{E}}$  will be called *the atoms of  $\mathcal{E}$* . It follows from the definition that in atomic event-systems, atoms are not decomposable, non-atoms are uniquely decomposable into atoms and events preserve atoms.

Since the set  $\mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$  is infinite, it is not possible to decide whether  $\mathcal{E}$  is atomic by exhaustively checking the connected component of every monomial in  $\mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$ . The following is sometimes helpful in deciding whether a finite-dimensional event-system is atomic (proof not provided).

Let  $\mathcal{E}$  be an event-system of dimension  $n$  with no event of the form  $\sigma + \tau N$ . Let  $B_{\mathcal{E}} = \{X_i \mid \text{For all } \sigma, \tau \in \mathbb{C} \setminus \{0\} \text{ and } N \in \mathbb{M}_{\infty}: \sigma X_i + \tau N \notin \mathcal{E}\}$ . Then  $\mathcal{E}$  is atomic iff there exist  $M_1 \in C_{\mathcal{E}}(X_1) \cap \mathbb{M}_{B_{\mathcal{E}}}$ ,  $M_2 \in C_{\mathcal{E}}(X_2) \cap \mathbb{M}_{B_{\mathcal{E}}}$ ,  $\dots$ ,  $M_n \in C_{\mathcal{E}}(X_n) \cap \mathbb{M}_{B_{\mathcal{E}}}$  such that:

$$\text{For all } \sigma \prod_{i=1}^n X_i^{a_i} - \tau \prod_{i=1}^n X_i^{b_i} \in \mathcal{E}, \quad \prod_{i=1}^n M_i^{a_i} = \prod_{i=1}^n M_i^{b_i}. \quad (1.1)$$

We have shown (proof not provided) that if  $\mathcal{E}$  and  $B_{\mathcal{E}}$  are as above, and there exist  $M_1 \in C_{\mathcal{E}}(X_1) \cap \mathbb{M}_{B_{\mathcal{E}}}$ ,  $M_2 \in C_{\mathcal{E}}(X_2) \cap \mathbb{M}_{B_{\mathcal{E}}}$ ,  $\dots$ ,  $M_n \in C_{\mathcal{E}}(X_n) \cap \mathbb{M}_{B_{\mathcal{E}}}$  and there exists  $\sigma \prod_{i=1}^n X_i^{a_i} - \tau \prod_{i=1}^n X_i^{b_i} \in \mathcal{E}$  such that  $\prod_{i=1}^n M_i^{a_i} \neq \prod_{i=1}^n M_i^{b_i}$ , then  $\mathcal{E}$  is not atomic. Hence, to check whether an event-system with no event of the form  $\sigma + \tau N$  is atomic, it suffices to examine an arbitrary choice of  $M_1 \in C_{\mathcal{E}}(X_1) \cap \mathbb{M}_{B_{\mathcal{E}}}$ ,  $M_2 \in C_{\mathcal{E}}(X_2) \cap \mathbb{M}_{B_{\mathcal{E}}}$ ,  $\dots$ ,  $M_n \in C_{\mathcal{E}}(X_n) \cap \mathbb{M}_{B_{\mathcal{E}}}$ , if one exists, and check whether (1.1) above holds.

*Example 1* Let  $\mathcal{E} = \{X_2^2 - X_1^2\}$ . Then  $B_{\mathcal{E}} = \{X_1, X_2\}$ . Let  $M_1 = X_1$  and  $M_2 = X_2$ . Trivially,  $M_1, M_2 \in \mathbb{M}_{B_{\mathcal{E}}}$ ,  $M_1 \in C_{\mathcal{E}}(X_1)$  and  $M_2 \in C_{\mathcal{E}}(X_2)$ . Consider the event  $X_2^2 - X_1^2$ . Since  $M_2^2 = X_2^2 \neq X_1^2 = M_1^2$ ,  $\mathcal{E}$  is not atomic. Note that the event  $X_2^2 - X_1^2$  does not preserve atoms.

*Example 2* Let  $\mathcal{E} = \{X_4^2 - X_2, X_5^2 - X_3, X_2X_3 - X_1\}$ . Then  $B_{\mathcal{E}} = \{X_4, X_5\}$ . Let  $M_1 = X_4^2X_5^2$ ,  $M_2 = X_4^2$ ,  $M_3 = X_5^2$ ,  $M_4 = X_4$ ,  $M_5 = X_5$ . Clearly these are all in  $\mathbb{M}_{B_{\mathcal{E}}}$ .  $X_5^2 - X_3 \in \mathcal{E}$  implies  $M_3 \in C_{\mathcal{E}}(X_3)$ .  $X_4^2 - X_2 \in \mathcal{E}$  implies  $M_2 \in C_{\mathcal{E}}(X_2)$ . Since  $(X_1, X_2X_3, X_2X_5^2, X_4^2X_5^2)$  is a path in  $G_{\mathcal{E}}$ , we have  $M_1 \in C_{\mathcal{E}}(X_1)$ . For the event  $X_4^2 - X_2$ , we have  $M_4^2 = X_4^2 = M_2$ . For the event  $X_5^2 - X_3$ , we have  $M_5^2 = X_5^2 = M_3$ . For the event  $X_2X_3 - X_1$ , we have  $M_2M_3 = X_4^2X_5^2 = M_1$ . Therefore,  $\mathcal{E}$  is atomic.

Note that it is possible to have an atomic event-system where  $A_{\mathcal{E}}$  is the empty set. For example:

*Example 3* Let  $\mathcal{E} = \{1 - X_1\}$ . In this case,  $S = \{X_1\}$  and  $\mathbb{M}_S$  is the set  $\{1, X_1, X_1^2, X_1^3, \dots\}$ . It is clear that  $\mathbb{M}_S$  forms a single connected component  $C$  in  $G_{\mathcal{E}}$ . Hence,  $X_1$  is not in  $A_{\mathcal{E}}$ , and  $A_{\mathcal{E}} = \emptyset$ . 1 is the only monomial in  $\mathbb{M}_{A_{\mathcal{E}}}$ . Since 1 is in  $C$ ,  $\mathcal{E}$  is atomic.

### 1.3 Finite Event-Systems

The study of infinite event-systems is embryonic and appears to be quite challenging. In the rest of this chapter only finite event-systems (i.e., where the set  $\mathcal{E}$  is finite) will be considered. It is clear that all finite event-systems are finite-dimensional.

**Definition 12** (*Stoichiometric matrix*) Let  $\mathcal{E} = \{e_1, e_2, \dots, e_m\}$  be an event-system of dimension  $n$ . Let  $i \leq n$  and  $j \leq m$  be positive integers. Let  $e_j = \sigma M + \tau N$ , where  $M \prec N$ . Then  $\gamma_{j,i}$  is the number of times  $X_i$  divides  $N$  minus the number of times  $X_i$  divides  $M$ . The *stoichiometric matrix*  $\Gamma_{\mathcal{E}}$  of  $\mathcal{E}$  is the  $m \times n$  matrix of integers  $\Gamma_{\mathcal{E}} = (\gamma_{j,i})_{m \times n}$ .

*Example 4* Let  $e_1 = 0.5X_2^5 - 500X_1X_2^3X_7$ . Let  $\mathcal{E} = \{e_1\}$ . Then  $\gamma_{1,1} = 1$ ,  $\gamma_{1,2} = -2$ ,  $\gamma_{1,7} = 1$  and for all other  $i$ ,  $\gamma_{1,i} = 0$ , hence  $\Gamma_{\mathcal{E}} = (1 \ -2 \ 0 \ 0 \ 0 \ 0 \ 1)$ .

**Definition 13** Let  $\mathcal{E} = \{e_1, \dots, e_m\}$  be a finite event-system of dimension  $n$ . Then:

1.  $P_{\mathcal{E}}$  is the column vector  $\langle P_1, P_2, \dots, P_n \rangle^T = \Gamma_{\mathcal{E}}^T \langle e_1, e_2, \dots, e_m \rangle^T$ .
2. Let  $\alpha \in \mathbb{C}^n$ . Then  $\alpha$  is an  $\mathcal{E}$ -equilibrium point iff for  $i = 1, 2, \dots, n$ :  $P_i(\alpha) = 0$ .

The  $P_i$ 's arise from the Law of Mass Action in chemistry. For a system of chemical reactions, the  $P_i$ 's are the right-hand sides of the differential equations that describe the concentration kinetics. Definition 13 extends the Law of Mass Action to arbitrary event-systems, and hence, arbitrary sets of binomials.

It follows from the definition that for finite event-systems, all strong equilibrium points are equilibrium points, but the converse need not be true.

*Example 5* Let  $e_1 = X_2 - X_1$  and  $e_2 = X_2 - 2X_1$ . Let  $\mathcal{E} = \{e_1, e_2\}$ . Then  $\Gamma_{\mathcal{E}} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$  and  $p_{\mathcal{E}} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} 2X_2 - 3X_1 \\ 3X_1 - 2X_2 \end{pmatrix}$ . Therefore  $(2, 3)$  is an  $\mathcal{E}$ -equilibrium point. Since  $e_1(2, 3) = 1$ ,  $(2, 3)$  is not a strong  $\mathcal{E}$ -equilibrium point.

*Example 6* Let  $e_1 = 6 - X_1X_2$  and  $e_2 = 2X_2^2 - 9X_1$ . Let  $\mathcal{E} = \{e_1, e_2\}$ . Then  $\Gamma_{\mathcal{E}} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}$  and  $p_{\mathcal{E}} = \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} = \begin{pmatrix} 6 - X_1X_2 + 2X_2^2 - 9X_1 \\ 6 - X_1X_2 - 4X_2^2 + 18X_1 \end{pmatrix}$ . The point  $(2, 3)$  is a strong equilibrium point because  $e_1(2, 3) = 0$  and  $e_2(2, 3) = 0$ . Since  $P_1(2, 3) = e_1(2, 3) + e_2(2, 3) = 0$  and  $P_2(2, 3) = e_1(2, 3) - 2e_2(2, 3) = 0$ , the point  $(2, 3)$  is also an equilibrium point.

The event-system in Example 5 is not natural, whereas the one in Example 6 is. In Theorem 8, it is shown that if  $\mathcal{E}$  is a finite, natural event-system then all positive  $\mathcal{E}$ -equilibrium points are strong  $\mathcal{E}$ -equilibrium points.

**Definition 14** (*Event-process*) Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ . Let  $\langle P_1, P_2, \dots, P_n \rangle^T = p_{\mathcal{E}}$ . Let  $\Omega \subseteq \mathbb{C}$  be a non-empty simply-connected open set. Let  $f = \langle f_1, f_2, \dots, f_n \rangle$  where for  $i = 1, 2, \dots, n$ ,  $f_i: \mathbb{C} \rightarrow \mathbb{C}$  is defined on  $\Omega$ . Then  $f$  is an  $\mathcal{E}$ -process on  $\Omega$  iff for  $i = 1, 2, \dots, n$ :

1.  $f'_i$  exists on  $\Omega$ .
2.  $f'_i = P_i \circ f$  on  $\Omega$ .

Note that  $\mathcal{E}$ -processes evolve through complex time, and hence generalize the idea of the time-evolution of concentrations in a system of chemical reactions.

Definition 14 immediately implies that if  $f = \langle f_1, f_2, \dots, f_n \rangle$  is an  $\mathcal{E}$ -process on  $\Omega$ , then for  $i = 1, 2, \dots, n$ ,  $f_i$  is holomorphic on  $\Omega$ . In particular, for each  $i$  and all  $\alpha \in \Omega$ , there is a power series around  $\alpha$  that agrees with  $f_i$  on a disk of non-zero radius.

Systems of chemical reactions sometimes obey certain conservation laws. For example, they may conserve mass, or the total number of each kind of atom. Event-systems also sometimes obey conservation laws.

**Definition 15** (*Conservation law, Linear conservation law*) Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ . A function  $g: \mathbb{C}^n \rightarrow \mathbb{C}$  is a *conservation law of  $\mathcal{E}$*  iff  $g$  is holomorphic on  $\mathbb{C}^n$ ,  $g(\langle 0, 0, \dots, 0 \rangle) = 0$  and  $\nabla g \cdot P_{\mathcal{E}}$  is identically zero on  $\mathbb{C}^n$ . If  $g$  is a conservation law of  $\mathcal{E}$  and  $g$  is linear (i.e.  $\forall c \in \mathbb{C}, \forall \alpha, \beta \in \mathbb{C}^n, g(c\alpha + \beta) = cg(\alpha) + g(\beta)$ ), then  $g$  is a *linear conservation law of  $\mathcal{E}$* .

The event-system described in Example 5 has a linear conservation law  $g(X_1, X_2) = X_1 + X_2$ . The next theorem shows that conservation laws of  $\mathcal{E}$  are dynamical invariants of  $\mathcal{E}$ -processes.

**Theorem 1** *For all finite event-systems  $\mathcal{E}$ , for all conservation laws  $g$  of  $\mathcal{E}$ , for all simply-connected open sets  $\Omega \subseteq \mathbb{C}$ , for all  $\mathcal{E}$ -processes  $f$  on  $\Omega$ , there exists  $k \in \mathbb{C}$  such that  $g \circ f - k$  is identically zero on  $\Omega$ .*

*Proof* Let  $n$  be the dimension of  $\mathcal{E}$ . Let  $\langle P_1, P_2, \dots, P_n \rangle^T = p_{\mathcal{E}}$ . For all  $t \in \Omega$ , by Definition 14, for  $i = 1, 2, \dots, n$ ,  $f_i(t)$  and  $f'_i(t)$  are defined. Further, by Definition 15,  $g$  is holomorphic on  $\mathbb{C}^n$ . Hence,  $g \circ f$  is holomorphic on  $\Omega$ . Therefore, by the chain rule,  $(g \circ f)'(t) = (\nabla g|_{f(t)}) \cdot \langle f'_1(t), f'_2(t), \dots, f'_n(t) \rangle$ . By Definition 14, for all  $t \in \Omega$ ,  $\langle f'_1(t), f'_2(t), \dots, f'_n(t) \rangle = \langle P_1(f(t)), P_2(f(t)), \dots, P_n(f(t)) \rangle$ . From these, it follows that  $(g \circ f)'(t) = (\nabla g \cdot P_{\mathcal{E}})(f(t))$ . But by Definition 15,  $\nabla g \cdot P_{\mathcal{E}}$  is identically zero. Hence, for all  $t \in \Omega$ ,  $(g \circ f)'(t) = 0$ . In addition,  $\Omega$  is a simply-connected open set. Therefore, by [2, Theorem 11], there exists  $k \in \mathbb{C}$  such that  $g \circ f - k$  is identically zero on  $\Omega$ .

The next theorem shows a way to derive linear conservation laws of an event-system from its stoichiometric matrix.

**Theorem 2** *Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ . For all  $V \in \ker \Gamma_{\mathcal{E}}$ ,  $V \cdot \langle X_1, \dots, X_n \rangle$  is a linear conservation law of  $\mathcal{E}$ .*

*Proof* Let  $\Gamma = \Gamma_{\mathcal{E}}$ , then  $\ker \Gamma$  is orthogonal to the image of  $\Gamma^T$ . By the definition of  $P = P_{\mathcal{E}}$ , for all  $w \in \mathbb{C}^n$ ,  $P(w)$  lies in the image of  $\Gamma^T$ . Hence, for all  $V \in \ker \Gamma$ , for all  $w \in \mathbb{C}^n$ ,  $V \cdot P(w) = 0$ . But  $V$  is the gradient of  $V \cdot \langle X_1, \dots, X_n \rangle$ . It now follows from Definition 15 that  $V \cdot \langle X_1, \dots, X_n \rangle$  is a linear conservation law of  $\mathcal{E}$ .

**Definition 16** (*Primitive conservation law*) Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ . For all  $V \in \ker \Gamma_{\mathcal{E}}$ , the linear conservation law  $V \cdot \langle X_1, X_2, \dots, X_n \rangle$  is a *primitive conservation law*.

We can show that in physical event-systems all linear conservation laws are primitive and, in natural event-systems, all conservation laws arise from the primitive ones.

**Definition 17** (*Conservation class, Positive conservation class*) Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ . A coset of  $(\ker \Gamma_{\mathcal{E}})^{\perp}$  is a *conservation class of  $\mathcal{E}$* . If a conservation class of  $\mathcal{E}$  contains a positive point, then the class is a *positive conservation class of  $\mathcal{E}$* .

Equivalently,  $\alpha, \beta \in \mathbb{C}^n$  are in the same conservation class if and only if they agree on all primitive conservation laws. Note that if  $H$  is a conservation class of  $\mathcal{E}$  then it is closed in  $\mathbb{C}^n$ . The following theorem shows that the name “conservation class” is appropriate.

**Theorem 3** *Let  $\mathcal{E}$  be a finite event-system. Let  $\Omega \subset \mathbb{C}$  be a simply-connected open set containing 0. Let  $f$  be an  $\mathcal{E}$ -process on  $\Omega$ . Let  $H$  be a conservation class of  $\mathcal{E}$  containing  $f(0)$ . Then for all  $t \in \Omega$ ,  $f(t) \in H$ .*

*Proof* Let  $\mathcal{E}, \Omega, f, H$  and  $t$  be as in the statement of this theorem. For all  $V \in \ker \Gamma_{\mathcal{E}}$ , the primitive conservation law  $V \cdot \langle X_1, X_2, \dots, X_n \rangle$  is a dynamical invariant of  $f$ , from Theorem 2 and Theorem 1. Hence,

$$V \cdot \langle f_1(0), f_2(0), \dots, f_n(0) \rangle = V \cdot \langle f_1(t), f_2(t), \dots, f_n(t) \rangle$$

That is,

$$V \cdot \langle f_1(0) - f_1(t), f_2(0) - f_2(t), \dots, f_n(0) - f_n(t) \rangle = 0$$

Hence,  $f(t) - f(0)$  is in  $(\ker \Gamma_{\mathcal{E}})^{\perp}$ . By Definition 17,  $f(t) \in H$ .

## 1.4 Finite Physical Event-Systems

In this section, we investigate finite, physical event-systems—a generalization of systems of chemical reactions.

It is widely believed that systems of chemical reactions that begin with positive (respectively, non-negative) concentrations will have positive (respectively, non-negative) concentrations at all future times. This property has been addressed mathematically in numerous papers [6, p. 6], [5, Remark 3.4], [3, Theorem 3.2], [14, Lemma 2.1]. The notion of “system of chemical reactions” varies between papers. Several papers have provided no proof, incomplete proofs or inadequate proofs that this property holds for their systems. Sontag [14, Lemma 2.1] provides a lovely



proof of this property for the systems he considers—zero deficiency reaction networks with one linkage class. We shall prove in Theorem 4 that the property holds for finite, physical event-systems. Finite, physical event-systems have a large intersection with the systems considered by Sontag, but each includes a large class of systems that the other does not. We remark that our methods of proof differ from Sontag’s, but it is possible that Sontag’s proof might be adaptable to our setting.

Lemma 4 and Lemma 10 are proved here because they apply to finite, physical event-systems. However, they are only invoked in subsequent sections. Lemma 4 relates  $\mathcal{E}$ -processes to solutions of ordinary differential equations over the reals. Lemma 10 establishes that if an  $\mathcal{E}$ -process defined on the positive reals starts at a real, non-negative point, then its  $\omega$ -limit set is invariant and contains only real, non-negative points.

The next lemma shows that if two  $\mathcal{E}$ -processes evaluate to the same real point on a real argument then they must agree and be real-valued on an open interval containing that argument. The proof exploits the fact that  $\mathcal{E}$ -processes are analytic, by considering their power series expansions.

**Lemma 1** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega, \Omega' \subseteq \mathbb{C}$  be open and simply-connected, let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$  and let  $g = \langle g_1, g_2, \dots, g_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega'$ . If  $t_0 \in \Omega \cap \Omega' \cap \mathbb{R}$  and  $f(t_0) \in \mathbb{R}^n$  and  $f(t_0) = g(t_0)$ , then there exists an open interval  $I \subseteq \mathbb{R}$  such that  $t_0 \in I$  and for all  $t \in I$ :*

1.  $f(t) = g(t)$ .
2. For  $i = 1, 2, \dots, n$ : if  $\sum_{j=0}^{\infty} c_j (z - t_0)^j$  is the Taylor series expansion of  $f_i$  at  $t_0$  then for all  $j \in \mathbb{Z}_{\geq 0}$ ,  $c_j \in \mathbb{R}$ .
3.  $f(t) \in \mathbb{R}^n$ .

*Proof* Let  $k \in \mathbb{Z}_{\geq 0}$ . By Definition 14,  $f$  and  $g$  are vectors of functions analytic at  $t_0$ . For  $i = 1, 2, \dots, n$ , let  $f_i^{(k)}$  be the  $k$ th derivative of  $f_i$  and let  $f^{(k)} = \langle f_1^{(k)}, f_2^{(k)}, \dots, f_n^{(k)} \rangle$ . Define  $g_i^{(k)}$  and  $g^{(k)}$  similarly. To prove 1, it is enough to show that for  $i = 1, 2, \dots, n$ ,  $f_i$  and  $g_i$  have the same Taylor series around  $t_0$ . Let  $V_0 = \langle X_1, X_2, \dots, X_n \rangle$ . Let  $V_k = \text{Jac}(V_{k-1})P_{\mathcal{E}}$  (recall that if  $H = \langle h_1(X_1, X_2, \dots, X_m), h_2(X_1, X_2, \dots, X_m), \dots, h_n(X_1, X_2, \dots, X_m) \rangle$  is a vector of functions in  $m$  variables then  $\text{Jac}(H)$  is the  $n \times m$  matrix  $(\frac{\partial h_i}{\partial x_j})$ , where  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, m$ ). Let  $\langle V_{k,1}, V_{k,2}, \dots, V_{k,n} \rangle = V_k$ . We claim that  $f^{(k)} = V_k \circ f$  on  $\Omega$  and  $g^{(k)} = V_k \circ g$  on  $\Omega'$  and for  $i = 1, 2, \dots, n$ ,  $V_{k,i} \in \mathbb{R}[X_1, X_2, \dots, X_n]$ . We prove the claim by induction on  $k$ . If  $k = 0$ , the proof is immediate. If  $k \geq 1$ , on  $\Omega$ :

$$\begin{aligned}
 f^{(k)} &= (f^{(k-1)})' \\
 &= (V_{k-1} \circ f)' \quad (\text{Inductive hypothesis}) \\
 &= (\text{Jac}(V_{k-1}) \circ f) f' \quad (\text{Chain-rule of derivation}) \\
 &= (\text{Jac}(V_{k-1}) \circ f)(P_{\mathcal{E}} \circ f) \quad (f \text{ is an } \mathcal{E}\text{-process})
 \end{aligned}$$

$$\begin{aligned}
&= (\text{Jac}(V_{k-1})P_{\mathcal{E}}) \circ f \\
&= V_k \circ f
\end{aligned}$$

By a similar argument, we conclude that  $g^{(k)} = V_k \circ g$  on  $\Omega'$ . By the inductive hypothesis,  $V_{k-1}$  is a vector of polynomials in  $\mathbb{R}[X_1, X_2, \dots, X_n]$ . It follows that  $\text{Jac}(V_{k-1})$  is an  $n \times n$  matrix of polynomials in  $\mathbb{R}[X_1, X_2, \dots, X_n]$ . Since  $\mathcal{E}$  is physical,  $P_{\mathcal{E}}$  is a vector of polynomials in  $\mathbb{R}[X_1, X_2, \dots, X_n]$ . Therefore,  $V_k = \text{Jac}(V_{k-1})P_{\mathcal{E}}$  is a vector of polynomials in  $\mathbb{R}[X_1, X_2, \dots, X_n]$ . This establishes the claim.

We have proved that  $f^{(k)} = V_k \circ f$  on  $\Omega$  and  $g^{(k)} = V_k \circ g$  on  $\Omega'$ . Since, by assumption,  $t_0 \in \Omega \cap \Omega'$  and  $f(t_0) = g(t_0)$ , it follows that  $f^{(k)}(t_0) = g^{(k)}(t_0)$ . Therefore, for  $i = 1, 2, \dots, n$ ,  $f_i$  and  $g_i$  have the same Taylor series around  $t_0$ . For  $i = 1, 2, \dots, n$ , let  $a_i$  be the radius of convergence of the Taylor series of  $f_i$  around  $t_0$ . Let  $r_f = \min_{i \in \{1, 2, \dots, n\}} a_i$ . Define  $r_g$  similarly. Let  $D \subseteq \Omega \cap \Omega'$  be some non-empty open disk centered at  $t_0$  with radius  $r \leq \min(r_f, r_g)$ . Since  $\Omega$  and  $\Omega'$  are open sets and  $t_0 \in \Omega \cap \Omega'$ , such a disk must exist. Letting  $I = (t_0 - r, t_0 + r)$  completes the proof of 1.

By assumption,  $f(t_0) \in \mathbb{R}^n$ , and we have proved that  $f^{(k)} = V_k \circ f$  and  $V_k$  is a vector of polynomials in  $\mathbb{R}[X_1, X_2, \dots, X_n]$ . It follows that  $f^{(k)}(t_0) \in \mathbb{R}^n$ . Therefore, for  $i = 1, 2, \dots, n$ , all coefficients in the Taylor series of  $f_i$  around  $t_0$  are real. It follows that  $f_i$  is real valued on  $I$ , completing the proof of 3.

The next lemma is a kind of uniqueness result. It shows that if two  $\mathcal{E}$ -processes evaluate to the same real point at 0 then they must agree and be real-valued on every open interval containing 0 where both are defined. The proof uses continuity to extend the result of Lemma 1.

**Lemma 2** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega, \Omega' \subseteq \mathbb{C}$  be open and simply-connected, let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$  and let  $g = \langle g_1, g_2, \dots, g_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega'$ . If  $0 \in \Omega \cap \Omega'$  and  $f(0) \in \mathbb{R}^n$  and  $f(0) = g(0)$ , then for all open intervals  $I \subseteq \Omega \cap \Omega' \cap \mathbb{R}$  such that  $0 \in I$ , for all  $t \in I$ ,  $f(t) = g(t)$  and  $f(t) \in \mathbb{R}^n$ .*

*Proof* Assume there exists an open interval  $I \subseteq \Omega \cap \Omega' \cap \mathbb{R}$  such that  $0 \in I$  and  $B = \{t \in I \mid f(t) \neq g(t) \text{ or } f(t) \notin \mathbb{R}^n\} \neq \emptyset$ . Let  $B_P = B \cap \mathbb{R}_{\geq 0}$  and let  $B_N = B \cap \mathbb{R}_{< 0}$ . Note that  $B = B_P \cup B_N$ , hence,  $B_P \neq \emptyset$  or  $B_N \neq \emptyset$ . Suppose  $B_P \neq \emptyset$  and let  $t_P = \inf(B_P)$ . By Lemma 1, there exists an  $\varepsilon \in \mathbb{R}_{> 0}$  such that  $(-\varepsilon, \varepsilon) \cap B = \emptyset$ . Hence,  $t_P \geq \varepsilon > 0$ . By definition of  $t_P$ , for all  $t \in [0, t_P)$ ,  $f(t) = g(t)$  and  $f(t) \in \mathbb{R}^n$ . Since  $f$  and  $g$  are analytic at  $t_P$ , they are continuous at  $t_P$ . Therefore,  $f(t_P) = g(t_P)$  and  $f(t_P) \in \mathbb{R}^n$ . By Lemma 1, there exists an  $\varepsilon' \in \mathbb{R}_{> 0}$  such that for all  $t \in (t_P - \varepsilon', t_P + \varepsilon')$ ,  $f(t) = g(t)$  and  $f(t) \in \mathbb{R}^n$ , contradicting  $t_P$  being the infimum of  $B_P$ . Therefore,  $B_P = \emptyset$ . Using a similar argument, we can prove that  $B_N = \emptyset$ . Therefore,  $B = \emptyset$ , and for all  $t \in I$ ,  $f(t) = g(t)$  and  $f(t) \in \mathbb{R}^n$ .

The next lemma is a convenient technical result that lets us ignore the choice of origin for the time variable.

**Lemma 3** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega, \tilde{\Omega} \subseteq \mathbb{C}$  be open and simply connected, let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$  and let  $\tilde{f} = \langle \tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n \rangle$  be an  $\mathcal{E}$ -process on  $\tilde{\Omega}$ . Let  $u \in \Omega$  and  $\tilde{u} \in \tilde{\Omega}$  and  $\alpha \in \mathbb{R}^n$ . Let  $I \subseteq \mathbb{R}$  be an open interval. If*

1.  $f(u) = \tilde{f}(\tilde{u}) = \alpha$  and
2.  $0 \in I$  and
3. for all  $s \in I$ ,  $u + s \in \Omega$  and  $\tilde{u} + s \in \tilde{\Omega}$

then for all  $t \in I$ ,  $f(u + t) = \tilde{f}(\tilde{u} + t)$ .

*Proof* Suppose  $f(u) = \tilde{f}(\tilde{u}) = \alpha \in \mathbb{R}^n$ . Let  $\Omega_u = \{z \in \mathbb{C} \mid u + z \in \Omega\}$  and  $\tilde{\Omega}_{\tilde{u}} = \{z \in \mathbb{C} \mid \tilde{u} + z \in \tilde{\Omega}\}$ . Let  $h = \langle h_1, h_2, \dots, h_n \rangle$  where for  $i = 1, 2, \dots, n$ ,  $h_i: \Omega_u \rightarrow \mathbb{C}$  is such that for all  $z \in \Omega_u$ ,  $h_i(z) = f_i(u + z)$  and let  $\tilde{h} = \langle \tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n \rangle$  where for  $i = 1, 2, \dots, n$ ,  $\tilde{h}_i: \tilde{\Omega}_{\tilde{u}} \rightarrow \mathbb{C}$  is such that for all  $z \in \tilde{\Omega}_{\tilde{u}}$ ,  $\tilde{h}_i(z) = \tilde{f}_i(\tilde{u} + z)$ . Since  $u + z$  is differentiable on  $\Omega_u$  and for  $i = 1, 2, \dots, n$ ,  $f_i$  is differentiable on  $\Omega$ , it follows that for  $i = 1, 2, \dots, n$ ,  $h_i$  is differentiable on  $\Omega_u$ . Further, for  $i = 1, 2, \dots, n$ , for all  $z \in \Omega_u$ ,  $h'_i(z) = f'_i(u + z) = p_{\mathcal{E}}(f_i(u + z)) = p_{\mathcal{E}}(h_i(z))$ , so  $h$  is an  $\mathcal{E}$ -process on  $\Omega_u$ . Similarly,  $\tilde{h}$  is an  $\mathcal{E}$ -process on  $\tilde{\Omega}_{\tilde{u}}$ . Note that  $0 \in \Omega_u \cap \tilde{\Omega}_{\tilde{u}}$  because  $u \in \Omega$  and  $\tilde{u} \in \tilde{\Omega}$  and that  $h(0) = \tilde{h}(0) = \alpha$  because  $f(u) = \tilde{f}(\tilde{u}) = \alpha$ . By Lemma 2, for all open intervals  $I \subseteq \Omega_u \cap \tilde{\Omega}_{\tilde{u}} \cap \mathbb{R}$  such that  $0 \in I$ , for all  $t \in I$ ,  $h(t) = \tilde{h}(t)$ , so  $f(u + t) = \tilde{f}(\tilde{u} + t)$ .

Because event-systems are defined over the complex numbers, we have access to results from complex analysis. However, there is a considerable body of results regarding ordinary differential equations over the reals. Definition 18 and Lemma 4 establish a relationship between  $\mathcal{E}$ -processes and solutions to systems of ordinary differential equations over the reals.

**Definition 18** (*Real event-process*) Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ . Let  $\langle P_1, P_2, \dots, P_n \rangle^T = p_{\mathcal{E}}$ . Let  $I \subseteq \mathbb{R}$  be an interval. Let  $h = \langle h_1, h_2, \dots, h_n \rangle$  where for  $i = 1, 2, \dots, n$ ,  $h_i: \mathbb{R} \rightarrow \mathbb{R}$  is defined on  $I$ . Then  $h$  is a real- $\mathcal{E}$ -process on  $I$  iff for  $i = 1, 2, \dots, n$ :

1.  $h'_i$  exists on  $I$ .
2.  $h'_i = P_i \circ h$  on  $I$ .

**Lemma 4** (All real- $\mathcal{E}$ -processes are restrictions of  $\mathcal{E}$ -processes) *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ . Let  $I \subseteq \mathbb{R}$  be an interval. Let  $h = \langle h_1, h_2, \dots, h_n \rangle$  be a real- $\mathcal{E}$ -process on  $I$ . Then there exist an open, simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $f$  on  $\Omega$  such that:*

1.  $I \subset \Omega$
2. For all  $t \in I$ :  $f(t) = h(t)$ .

*Proof* Let  $P = \langle P_1, P_2, \dots, P_n \rangle = p_{\mathcal{E}}$ . For  $i = 1, 2, \dots, n$ ,  $P_i$  is a polynomial and therefore analytic on  $\mathbb{C}^n$ . By Cauchy's existence theorem for ordinary differential equations with analytic right-hand sides [12], for all  $a \in I$ , there exist a non-empty open disk  $D_a \subseteq \mathbb{C}$  centered at  $a$  and functions  $f_{a,1}, f_{a,2}, \dots, f_{a,n}$  analytic on  $D_a$  such that for  $i = 1, 2, \dots, n$ :

1.  $f_{a,i}(a) = h_i(a)$
2.  $f'_{a,i}$  exists on  $D_a$  and for all  $t \in D_a$ :  $f'_{a,i}(t) = P_i(f_{a,1}(t), f_{a,2}(t), \dots, f_{a,n}(t))$ .  
That is,  $f_a = \langle f_{a,1}, f_{a,2}, \dots, f_{a,n} \rangle$  is an  $\mathcal{E}$ -process on  $D_a$ .

**Claim** For all  $a \in I$ , there exists  $\delta_a \in \mathbb{R}_{>0}$  such that for all  $t \in I \cap (a - \delta_a, a + \delta_a)$ :  $f_a(t) = h(t)$ . To see this, by Lemma 1, for all  $a \in I$  there exists  $\beta_a \in \mathbb{R}_{>0}$  such that for all  $t \in (a - \beta_a, a + \beta_a) \cap D_a$ ,  $f_a(t) \in \mathbb{R}^n$ . Let  $I_a = (a - \beta_a, a + \beta_a) \cap D_a$ . Note that  $f_a|_{I_a}$  is a real- $\mathcal{E}$ -process on  $I_a$ . By the theorem of uniqueness of solutions to differential equations with  $\mathcal{C}^1$  right-hand sides [8], there exists  $\gamma_a \in \mathbb{R}_{>0}$  such that for all  $t \in (a - \gamma_a, a + \gamma_a) \cap I_a \cap I$ ,  $f_a(t) = h(t)$ . Clearly, we can choose  $\delta_a \in \mathbb{R}_{>0}$  such that  $(a - \delta_a, a + \delta_a) \subseteq (a - \gamma_a, a + \gamma_a) \cap I_a$ . This establishes the claim.

For all  $a \in I$ , let  $\delta_a \in \mathbb{R}_{>0}$  be such that for all  $t \in I \cap (a - \delta_a, a + \delta_a)$ :  $f_a(t) = h(t)$ . Let  $\widehat{D}_a$  be an open disk centered at  $a$  of radius  $\delta_a$ .

**Claim** For all  $a_1, a_2 \in I$ , for all  $t \in \widehat{D}_{a_1} \cap \widehat{D}_{a_2}$ :  $f_{a_1}(t) = f_{a_2}(t)$ . To see this, suppose  $\widehat{D}_{a_1} \cap \widehat{D}_{a_2} \neq \emptyset$ . Let  $J = \widehat{D}_{a_1} \cap \widehat{D}_{a_2} \cap \mathbb{R}$ . Since  $\widehat{D}_{a_1}$  and  $\widehat{D}_{a_2}$  are open disks centered on the real line,  $J$  is a non-empty open real interval. For all  $t \in J$ , by the claim above,  $f_{a_1}(t) = h(t)$  and  $f_{a_2}(t) = h(t)$ . Hence,  $f_{a_1}(t) = f_{a_2}(t)$ . Since  $J$  is a non-empty interval,  $J$  contains an accumulation point. Since  $f_{a_1}$  and  $f_{a_2}$  are analytic on  $\widehat{D}_{a_1} \cap \widehat{D}_{a_2}$  and  $\widehat{D}_{a_1} \cap \widehat{D}_{a_2}$  is simply connected, for all  $t \in \widehat{D}_{a_1} \cap \widehat{D}_{a_2}$ :  $f_{a_1}(t) = f_{a_2}(t)$ . This establishes the claim.

Let  $\Omega = \bigcup_{a \in I} \widehat{D}_a$ . Clearly,  $I \subset \Omega$ .  $\Omega$  is a union of open discs, and is therefore open.

For all  $t \in \Omega$ , there exists  $a \in I$  such that  $t \in \widehat{D}_a$ . Since  $\widehat{D}_a$  is a disk,  $t$  and  $a$  are path-connected in  $\Omega$ . Since  $I$  is path-connected, and  $I \subseteq \Omega$ , it follows that  $\Omega$  is path-connected.

To see that  $\Omega$  is simply-connected, consider the function  $R: [0, 1] \times \Omega \rightarrow \Omega$  given by  $(u, z) \mapsto \operatorname{Re}(z) + i \operatorname{Im}(z)(1-u)$ . Observe that  $R$  is continuous on  $[0, 1] \times \Omega$ , and for all  $z \in \Omega$ :  $R(0, z) = z$ ,  $R(1, \Omega) \subset \Omega$ , and for all  $u \in [0, 1]$ , for all  $z \in \Omega \cap \mathbb{R}$ :  $R(u, z) \in \Omega$ . Therefore,  $R$  is a deformation retraction. Note that  $R(0, \Omega) = \Omega$  and  $R(1, \Omega) \subseteq \mathbb{R}$ , and  $\Omega$  is path-connected together imply that  $R(1, \Omega)$  is a real interval. Hence,  $R(1, \Omega)$  is simply-connected. Since  $R$  was a deformation retraction,  $\Omega$  is simply-connected.

Let  $f: \Omega \rightarrow \mathbb{C}^n$  be the unique function such that for all  $a \in I$ , for all  $t \in \widehat{D}_a$ :  $f(t) = f_a(t)$ . By the claim above and from the definition of  $\Omega$ ,  $f$  is well-defined.

Observe that for all  $t \in I$ ,

$$\begin{aligned} h(t) &= f_t(t) \quad (\text{Definition of } f_t) \\ &= f(t) \quad (I \subset \Omega \text{ and definition of } f). \end{aligned}$$

**Claim**  $f$  is an  $\mathcal{E}$ -process on  $\Omega$ . From the definitions of  $\Omega$  and  $f$ , for all  $t \in \Omega$ , there exists  $a \in I$  such that  $t \in \widehat{D}_a$  and for all  $s \in \widehat{D}_a$ ,  $f(s) = f_a(s)$ . Since  $f_a$  is an  $\mathcal{E}$ -process on  $\widehat{D}_a$ , the claim follows.

In Theorem 4, we prove that if  $\mathcal{E}$  is a finite, physical event-system, then  $\mathcal{E}$ -processes that begin at positive (respectively non-negative) points remain positive (respectively non-negative) through all forward real time where they are defined. In fact, Theorem 4 establishes more detail about  $\mathcal{E}$ -processes. In particular, if at some time a species' concentration is positive, then it will be positive at subsequent times.

**Theorem 4** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be open and simply-connected, and let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$ . If  $I \subseteq \Omega \cap \mathbb{R}_{\geq 0}$  is connected and  $0 \in I$  and  $f(0)$  is a non-negative point then for  $k = 1, 2, \dots, n$  either:*

1. For all  $t \in I$ ,  $f_k(t) = 0$ , or
2. For all  $t \in I \cap \mathbb{R}_{>0}$ ,  $f_k(t) \in \mathbb{R}_{>0}$ .

The proof of Theorem 4 is highly technical, and relies on a detailed examination of the vector of polynomials  $p_{\mathcal{E}}$ . This allows us to show (Lemma 7) that if  $f = \langle f_1, f_2, \dots, f_n \rangle$  is an  $\mathcal{E}$ -process that at real time  $t_0$  is non-negative, then each  $f_i$  is “right non-negative.” That is, the Taylor series expansion of  $f_i$  around  $t_0$  has real coefficients and the first non-zero coefficient, if any, is positive. Further, (Lemma 9) if  $f_i(t_0) = 0$  and its Taylor series expansion has a non-zero coefficient, then there exists  $k$  such that  $f_k(t_0) = 0$  and the first derivative of  $f_k$  with respect to time is positive at  $t_0$ .

**Definition 19** Let  $n \in \mathbb{Z}_{>0}$  and let  $k \in \{1, 2, \dots, n\}$ . A polynomial  $f \in \mathbb{R}[X_1, X_2, \dots, X_n]$  is non-nullifying with respect to  $k$  iff there exist  $m \in \mathbb{N}$ ,  $c_1, c_2, \dots, c_m \in \mathbb{R}_{>0}$ ,  $M_1, M_2, \dots, M_m \in \mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$  and  $h \in \mathbb{R}[X_1, X_2, \dots, X_n]$  such that  $f = \sum_{i=1}^m c_i M_i + X_k h$ .

Observe that for all  $k$ , the polynomial 0 is non-nullifying with respect to  $k$ .

**Lemma 5** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ . Let  $\langle P_1, P_2, \dots, P_n \rangle = p_{\mathcal{E}}$ . Then, for all  $i \in \{1, 2, \dots, n\}$ ,  $P_i$  is non-nullifying with respect to  $i$ .*

*Proof* Let  $m = |\mathcal{E}|$ . Let  $(\gamma_{j,i})_{m \times n} = \Gamma_{\mathcal{E}}$ . Since  $\mathcal{E}$  is physical, there exist  $\sigma_1, \sigma_2, \dots, \sigma_m, \tau_1, \tau_2, \dots, \tau_m \in \mathbb{R}_{>0}$  and  $M_1, M_2, \dots, M_m, N_1, N_2, \dots, N_m \in \mathbb{M}_{\infty}$  such that for  $j = 1, 2, \dots, m$ :  $M_j \prec N_j$  and  $\{\sigma_1 M_1 - \tau_1 N_1, \sigma_2 M_2 - \tau_2 N_2, \dots, \sigma_m M_m - \tau_m N_m\} = \mathcal{E}$ . Let  $i \in \{1, 2, \dots, n\}$ .

From the definition of  $p_{\mathcal{E}}$ ,  $P_i = \sum_{j=1}^m \gamma_{j,i} (\sigma_j M_j - \tau_j N_j)$ . It is sufficient to prove that for  $j = 1, 2, \dots, m$ :  $\gamma_{j,i} (\sigma_j M_j - \tau_j N_j)$  is non-nullifying with respect to  $i$ . Let  $j \in \{1, 2, \dots, m\}$ . If  $\gamma_{j,i} = 0$  then  $\gamma_{j,i} (\sigma_j M_j - \tau_j N_j) = 0$  which is non-nullifying with respect to  $i$ . If  $\gamma_{j,i} > 0$  then, from the definition of  $\Gamma_{\mathcal{E}}$ ,  $X_i \mid N_j$  and

$$\gamma_{j,i} (\sigma_j M_j - \tau_j N_j) = \gamma_{j,i} \sigma_j M_j + X_i \left( -\gamma_{j,i} \tau_j \frac{N_j}{X_i} \right)$$

which is non-nullifying with respect to  $i$  since  $\gamma_{j,i} \sigma_j > 0$ . Similarly, if  $\gamma_{j,i} < 0$  then  $X_i \mid M_j$  and

$$\gamma_{j,i}(\sigma_j M_j - \tau_j N_j) = -\gamma_{j,i}\tau_j N_j + X_i \gamma_{j,i} \sigma_j \frac{M_j}{X_i}$$

which is non-nullifying with respect to  $i$  since  $-\gamma_{j,i}\tau_j > 0$ . Hence,  $P_i$  is non-nullifying with respect to  $i$ .

**Definition 20** Let  $t_0 \in \mathbb{C}$ , let  $f: \mathbb{C} \rightarrow \mathbb{C}$  be analytic at  $t_0$  and let  $f(t) = \sum_{k=0}^{\infty} c_k(t - t_0)^k$  be the Taylor series expansion of  $f$  around  $t_0$ . Then  $O(f, t_0)$  is the least  $k$  such that  $c_k \neq 0$ . If for all  $k$ ,  $c_k = 0$ , then  $O(f, t_0) = \infty$ .

**Definition 21** (*Right non-negative*) Let  $t_0 \in \mathbb{R}$ , let  $f: \mathbb{C} \rightarrow \mathbb{C}$  be analytic at  $t_0$  and let  $f(t) = \sum_{k=0}^{\infty} c_k(t - t_0)^k$  be the Taylor series expansion of  $f$  around  $t_0$ . Then  $f$  is RNN at  $t_0$  iff both:

1. For all  $k \in \mathbb{N}$ ,  $c_k \in \mathbb{R}$  and
2. Either  $O(f, t_0) = \infty$  or  $c_{O(f,t_0)} \in \mathbb{R}_{>0}$ .

**Lemma 6** Let  $t_0 \in \mathbb{C}$ . Let  $f, g: \mathbb{C} \rightarrow \mathbb{C}$  be functions analytic at  $t_0$ . Then:

1.  $O(f \cdot g, t_0) = O(f, t_0) + O(g, t_0)$ .
2. If  $t_0 \in \mathbb{R}$  and  $f, g$  are RNN at  $t_0$  then  $f \cdot g$  is RNN at  $t_0$ .

The proof is obvious.

**Lemma 7** Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be open and simply-connected and let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$ . For all  $t_0 \in \Omega \cap \mathbb{R}$ , if  $f(t_0) \in \mathbb{R}_{\geq 0}^n$  then for  $i = 1, 2, \dots, n$ :  $f_i$  is RNN at  $t_0$ .

*Proof* Suppose  $t_0 \in \Omega \cap \mathbb{R}$  and  $f(t_0) \in \mathbb{R}_{\geq 0}^n$ . Let  $p = \langle P_1, P_2, \dots, P_n \rangle = p_{\mathcal{E}}$ . Let  $C = \{i \mid f_i \text{ is not RNN at } t_0\}$ .

For the sake of contradiction, suppose  $C \neq \emptyset$ . Let  $m = \min_{i \in C} O(f_i, t_0)$ . Let  $k \in C$  be such that  $O(f_k, t_0) = m$ . Let  $f_k(t) = \sum_{i=0}^{\infty} a_i(t - t_0)^i$  be the Taylor series expansion of  $f_k$  around  $t_0$ . Since  $\mathcal{E}$  is physical and  $t_0 \in \mathbb{R}$  and  $f(t_0) \in \mathbb{R}_{\geq 0}^n$ , it follows from Lemma 1.2 that for all  $i \in \mathbb{N}$ ,  $a_i \in \mathbb{R}$ . Further:

$$a_0 = a_1 = \dots = a_{m-1} = 0 \quad (O(f_k, t_0) = m.) \quad (1.2)$$

$$a_m \in \mathbb{R}_{<0} \quad (f_k \text{ is not RNN at } t_0.) \quad (1.3)$$

Since  $f(t_0) \in \mathbb{R}_{\geq 0}^n$  and  $a_m \in \mathbb{R}_{<0}$  and  $a_0 = f_k(t_0)$ , it follows that  $m > 0$ .

Consider  $f'_k = P_k \circ f$ . By differentiation, the Taylor series expansion of  $f'_k$  at  $t_0$  is:

$$f'_k(t) = \sum_{i=0}^{\infty} (i+1)a_{i+1}(t - t_0)^i. \quad (1.4)$$

From Lemma 5,  $P_k$  is non-nullifying. Hence, there exist  $l \in \mathbb{N}$ ,  $b_1, b_2, \dots, b_l \in \mathbb{R}_{>0}$ ,  $M_1, M_2, \dots, M_l \in \mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$  and  $h \in \mathbb{R}[X_1, X_2, \dots, X_n]$  such that  $P_k = \sum_{j=1}^l b_j M_j + X_k \cdot h$ . Then for all  $t \in \Omega$ :

$$f'_k(t) = P_k \circ f(t) = \sum_{j=1}^l b_j M_j \circ f(t) + f_k(t) \cdot (h \circ f(t)) \quad (1.5)$$

Since  $h$  is a polynomial,  $h \circ f$  is analytic at  $t_0$ . Therefore,  $f_k \cdot (h \circ f)$  is analytic at  $t_0$ . Let  $\sum_{i=0}^{\infty} c_i (t-t_0)^i$  be the Taylor series expansion of  $f_k \cdot (h \circ f)$  at  $t_0$ . Similarly, for  $j = 1, 2, \dots, l$ ,  $b_j M_j \circ f$  is analytic at  $t_0$ . Let  $\sum_{i=0}^{\infty} d_{j,i} (t-t_0)^i$  be the Taylor series expansion of  $b_j M_j \circ f$  at  $t_0$ . From (1.4), (1.5), equating Taylor series coefficients, for  $i = 0, 1, \dots, m-1$ :

$$(i+1)a_{i+1} = c_i + \sum_{j=1}^l d_{j,i} \quad (1.6)$$

From Lemma 6.1,

$$O(f_k \cdot (h \circ f), t_0) = O(f_k, t_0) + O(h \circ f, t_0) \geq O(f_k, t_0) = m$$

Hence,

$$c_0 = c_1 = \dots = c_{m-1} = 0. \quad (1.7)$$

From (1.2), (1.6), (1.7), for  $i = 0, 1, \dots, m-2$ :

$$\sum_{j=1}^l d_{j,i} = 0 \quad (1.8)$$

Since  $m > 0$ , from (1.3), (1.6), (1.7):

$$\sum_{j=1}^l d_{j,m-1} = ma_m \in \mathbb{R}_{<0} \quad (1.9)$$

Let  $i_0 = \min_{j=1,2,\dots,l} \{O(b_j M_j \circ f, t_0)\}$ . From (1.9), it follows that  $i_0 \leq m-1$ .  
 Case 1: For  $j = 1, 2, \dots, l$ :  $d_{j,i_0} \in \mathbb{R}_{\geq 0}$ . From the definition of  $i_0$  it follows that  $\sum_{j=1}^l d_{j,i_0} \in \mathbb{R}_{>0}$ . If  $i_0 < m-1$ , this contradicts (1.8). If  $i_0 = m-1$ , this contradicts (1.9).

Case 2: There exists  $j_0 \in \{1, 2, \dots, l\}$  such that  $d_{j_0,i_0} \in \mathbb{R}_{<0}$ . From the definition of  $i_0$ ,  $O(b_{j_0} M_{j_0}, t_0) = i_0 \leq m-1$ . Therefore, for each  $i$  such that  $X_i | M_{j_0}$ ,  $O(f_i, t_0) \leq m-1$ . From the definitions of  $C$  and  $m$ , this implies that for each  $i$  such that  $X_i | M_{j_0}$ ,  $f_i$  is RNN at  $t_0$ . Since  $b_{j_0} \in \mathbb{R}_{>0}$ , it follows that  $b_{j_0} M_{j_0} \circ f$  is a product of RNN functions. Hence, by Lemma 6.2,  $b_{j_0} M_{j_0} \circ f$  is RNN at  $t_0$  and  $d_{j_0,i_0} \in \mathbb{R}_{>0}$ , a contradiction.

Hence, for  $i = 1, 2, \dots, n$ ,  $f_i$  is RNN at  $t_0$ .

**Lemma 8** *Let  $t_0 \in \mathbb{R}$  and let  $f$  be a function RNN at  $t_0$ . There exists an  $\varepsilon \in \mathbb{R}_{>0}$  such that either for all  $t \in (t_0, t_0 + \varepsilon)$ ,  $f(t) \in \mathbb{R}_{>0}$  or for all  $t \in (t_0, t_0 + \varepsilon)$ ,  $f(t) = 0$ .*

*Proof* Let  $m = O(f, t_0)$ . If  $m = \infty$ ,  $f$  is identically zero and the lemma follows immediately. Otherwise, let  $f^{(m)}$  denote the  $m$ th derivative of  $f$ . Since  $f$  is RNN at  $t_0$  and has order  $m$ ,  $f^{(m)}(t_0) \in \mathbb{R}_{>0}$ . Since  $f$  is analytic at  $t_0$ ,  $f^{(m)}$  is analytic at  $t_0$ , and hence continuous at  $t_0$ . By continuity, there exists  $\varepsilon \in \mathbb{R}_{>0}$  such that for all  $\tau \in [t_0, t_0 + \varepsilon]$ :  $f^{(m)}(\tau) \in \mathbb{R}_{>0}$ . From Taylor's theorem, for all  $t \in (t_0, t_0 + \varepsilon)$ , there exists  $\tau \in [t_0, t_0 + \varepsilon]$  such that:

$$f(t) = \frac{(t - t_0)^m}{m!} f^{(m)}(\tau)$$

Therefore,  $f(t) \in \mathbb{R}_{>0}$ .

Note that Lemmas 7 and 8 together already imply that if  $\mathcal{E}$  is a finite, physical event-system, then  $\mathcal{E}$ -processes that begin at non-negative points remain non-negative through all forward real time where they are defined. This result is weaker than Theorem 4.

**Lemma 9** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be open and simply-connected, let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$ . Let  $t_0 \in \Omega$ . If  $f(t_0)$  is non-negative and there exists  $j \in \{1, 2, \dots, n\}$  such that  $0 < O(f_j, t_0) < \infty$  then there exists  $k \in \{1, 2, \dots, n\}$  such that  $O(f_k, t_0) = 1$ .*

*Proof* Suppose  $f(t_0) \in \mathbb{R}_{\geq 0}^n$ . Let  $C = \{i | 0 < O(f_i, t_0) < \infty\}$ . Suppose  $C \neq \emptyset$ . Let  $m = \min_{i \in C} O(f_i, t_0)$ . There exists  $k \in C$  such that  $O(f_k, t_0) = m$ .

Let  $p = \langle P_1, P_2, \dots, P_n \rangle = p_{\mathcal{E}}$ . From Lemma 5,  $P_k$  is non-nullifying with respect to  $k$ . Hence, there exist  $l \in \mathbb{N}$ ,  $b_1, b_2, \dots, b_l \in \mathbb{R}_{>0}$ ,  $M_1, M_2, \dots, M_l \in \mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$  and  $h \in \mathbb{R}[X_1, X_2, \dots, X_n]$  such that  $P_k = \sum_{j=1}^l b_j M_j + X_k \cdot h$ .

For all  $t \in \Omega$ :  $f'_k(t) = P_k \circ f(t) = \sum_{j=1}^l b_j M_j \circ f(t) + f_k(t) \cdot (h \circ f(t))$ . From Lemma 6.1,  $O(f'_k \cdot (h \circ f), t_0) = O(f_k, t_0) + O(h \circ f, t_0) \geq O(f_k, t_0) = m$ . It follows that:

$$m - 1 = O(f'_k, t_0) = O\left(\sum_{j=1}^l b_j M_j \circ f, t_0\right) \quad (1.10)$$

From Lemmas (6.2) and Lemma 7, for  $j = 1, 2, \dots, l$ :  $b_j M_j \circ f$  is RNN at  $t_0$ . It follows that  $O(\sum_{j=1}^l b_j M_j \circ f, t_0) = \min_{j=1, 2, \dots, l} O(b_j M_j \circ f, t_0)$ . From Eq. (1.10),  $m - 1 = \min_{j=1, 2, \dots, l} O(b_j M_j \circ f, t_0)$ . Hence, there exists  $j_0$  such that  $O(b_{j_0} M_{j_0} \circ f, t_0) = m - 1$ . From Lemma (6.1), for all  $i$  such that  $X_i | M_{j_0}$ ,  $O(f_i, t_0) \leq m - 1$ . From the definition of  $m$ , for all  $i$  such that  $X_i | M_{j_0}$ ,  $O(f_i, t_0) = 0$ . It follows that  $m - 1 = O(b_{j_0} M_{j_0} \circ f, t_0) = 0$ . Hence,  $m = 1$ .

We are now ready to prove Theorem 4.



*Proof* (Proof of Theorem 4) Suppose  $I \subseteq \Omega \cap \mathbb{R}_{\geq 0}$  is connected and  $0 \in I$  and  $f(0)$  is a non-negative point. If  $I \cap \mathbb{R}_{> 0} = \emptyset$ , the theorem is immediate. Suppose  $I \cap \mathbb{R}_{> 0} \neq \emptyset$ .

It is clear that for all  $k$ ,  $O(f_k, 0) = \infty$  iff for all  $t \in I$ ,  $f_k(t) = 0$ . Let  $C = \{i \mid O(f_i, 0) \neq \infty\}$ . From Lemmas 7 and 8, for all  $k \in C$ , there exists  $\varepsilon_k \in I \cap \mathbb{R}_{> 0}$  such that for all  $t \in (0, \varepsilon_k)$ :  $f_k(t) \in \mathbb{R}_{> 0}$ .

Suppose for the sake of contradiction that there exist  $i \in C$  and  $t \in I \cap \mathbb{R}_{> 0}$  such that  $f_i(t) \notin \mathbb{R}_{> 0}$ . From Lemma 2,  $f_i(t) \in \mathbb{R}$ . Since  $f_i(\varepsilon_i/2) \in \mathbb{R}_{> 0}$  and  $f_i(t) \in \mathbb{R}_{\leq 0}$ , by continuity there exists  $t' \in I \cap \mathbb{R}_{> 0}$  such that  $f_i(t') = 0$ .

Let  $t_0 = \inf\{t \in I \cap \mathbb{R}_{> 0} \mid \text{There exists } i \in C \text{ with } f_i(t) = 0\}$ . It follows that:

1.  $t_0 \in \mathbb{R}_{> 0}$  because  $t_0 \geq \min_{i \in C} \{\varepsilon_i\}$ .
2.  $f(t_0) \in \mathbb{R}_{\geq 0}^n$ , from the definition of  $t_0$ .
3. There exists  $i_1 \in C$  such that  $O(f_{i_1}, t_0) = 1$ . This follows because there exist  $i_0 \in C$  and  $T \subseteq I \cap \mathbb{R}_{> 0}$  such that  $t_0 = \inf(T)$  and for all  $t \in T$ :  $f_{i_0}(t) = 0$ . By continuity,  $f_{i_0}(t_0) = 0$ . Hence,  $O(f_{i_0}, t_0) > 0$ . Since  $i_0 \in C$ ,  $O(f_{i_0}, 0) \neq \infty$ . By connectedness of  $I$ ,  $O(f_{i_0}, t_0) \neq \infty$ . Therefore,  $0 < O(f_{i_0}, t_0) < \infty$ . Since  $f(t_0) \in \mathbb{R}_{\geq 0}^n$ , by Lemma (9), there exists  $i_1 \in \{1, 2, \dots, n\}$  such that  $O(f_{i_1}, t_0) = 1$ . Assume  $i_1 \notin C$ . Then  $O(f_{i_1}, 0) = \infty$ . By connectedness of  $I$ ,  $O(f_{i_1}, t_0) = \infty$ , contradicting that  $O(f_{i_1}, t_0) = 1$ . Hence,  $i_1 \in C$ . Hence,  $f_{i_1}(t_0) = 0$ . Since  $f(t_0) \in \mathbb{R}_{\geq 0}^n$ , by Lemma (7)  $f'_{i_1}(t_0) \in \mathbb{R}_{> 0}$ .

From the definition of  $t_0$ , for all  $t \in (0, t_0)$ ,  $f_{i_1}(t) \in \mathbb{R}_{> 0}$ . Since  $t_0 \in \mathbb{R}_{> 0}$ ,

$$f'_{i_1}(t_0) = \lim_{h \rightarrow 0^+} \frac{f_{i_1}(t_0) - f_{i_1}(t_0 - h)}{h} = \lim_{h \rightarrow 0^+} \frac{-f_{i_1}(t_0 - h)}{h} \in \mathbb{R}_{\leq 0},$$

a contradiction. The theorem follows.

There is a notion in chemistry that, for systems of chemical reactions, concentrations evolve through time to reach equilibrium. In later sections of this chapter, we will investigate this notion. In the remainder of this section of the chapter, we will prepare for that investigation.

**Definition 22** Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be open, simply connected and such that  $\mathbb{R}_{\geq 0} \subseteq \Omega$ , let  $f$  be an  $\mathcal{E}$ -process on  $\Omega$ , and let  $q \in \mathbb{C}^n$ . Then  $q$  is an  $\omega$ -limit point of  $f$  iff for all  $\varepsilon \in \mathbb{R}_{> 0}$  there exists a sequence of non-negative reals  $\{t_i\}_{i \in \mathbb{Z}_{> 0}}$  such that  $t_i \rightarrow \infty$  as  $i \rightarrow \infty$  and for all  $i \in \mathbb{Z}_{> 0}$ ,  $\|f(t_i) - q\|_2 < \varepsilon$ .

Sometimes, an  $\omega$ -limit is defined by the existence of a single sequence of times such that the value approaches the limit. The above definition is easily seen to be equivalent.

**Definition 23** Let  $\mathcal{E}$  be a finite event-system of dimension  $n$  and let  $S \subseteq \mathbb{C}^n$ .  $S$  is an *invariant set* of  $\mathcal{E}$  iff for all  $q \in S$ , for all open, simply-connected  $\Omega \subseteq \mathbb{C}$ , for all  $\mathcal{E}$ -processes  $f$  on  $\Omega$ , if  $0 \in \Omega$  and  $f(0) = q$  then for all  $t \in \mathbb{R}_{\geq 0}$  such that  $[0, t] \subseteq \Omega$ ,  $f(t) \in S$ .

**Lemma 10** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be open and simply connected, and let  $f$  be an  $\mathcal{E}$ -process on  $\Omega$ . If  $\mathbb{R}_{\geq 0} \subseteq \Omega$  and  $f(0)$  is a non-negative point, then the set of all  $\omega$ -limit points of  $f$  is an invariant set of  $\mathcal{E}$  and is contained in  $\mathbb{R}_{\geq 0}^n$ .*

*Proof* Let  $S$  be the set of all  $\omega$ -limit points of  $f$ . By Lemma 4, for all  $t \in \mathbb{R}_{\geq 0}$ ,  $f(t) \in \mathbb{R}_{\geq 0}^n$ , hence  $S \subseteq \mathbb{R}_{\geq 0}^n$ .

Let  $q \in S$ , let  $\tilde{\Omega} \subseteq \mathbb{C}$  be open, simply-connected, and such that  $0 \in \tilde{\Omega}$ , and let  $h$  be an  $\mathcal{E}$ -process on  $\tilde{\Omega}$  such that  $h(0) = q$ . Suppose  $u \in \mathbb{R}_{\geq 0}$  and  $[0, u] \subseteq \tilde{\Omega}$ . Since  $\mathcal{E}$  is finite and physical,  $p_{\mathcal{E}}|_{\mathbb{R}^n}$  can be viewed as a map  $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$  of class  $\mathcal{C}^1$ . By Lemma 2, for all  $t \in [0, u]$ ,  $h(t) \in \mathbb{R}^n$ , so  $h|_{[0, u]}$  can be viewed as a map  $X: [0, u] \rightarrow \mathbb{R}^n$  such that  $X' = f(X)$ . By Hirsch [8, p. 147], there exists a neighborhood  $U \subset \mathbb{R}^n$  of  $q$  and a constant  $K$  such that for all  $\alpha \in U$ , there exists a unique real- $\mathcal{E}$ -process  $\rho_{\alpha}$  defined on  $[0, u]$  with  $\rho_{\alpha}(0) = \alpha$  and  $\|\rho_{\alpha}(u) - h(u)\|_2 \leq K \|\alpha - q\|_2 \exp(Ku)$ . Observe that necessarily  $K \in \mathbb{R}_{\geq 0}$ . By Lemma 4 for all  $\alpha \in U$  there exists an open, simply-connected  $\Omega_{\alpha} \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $\rho_{\alpha}$  on  $\Omega_{\alpha}$  such that  $[0, u] \subseteq \Omega_{\alpha}$  and for all  $t \in [0, u]$ ,  $\rho_{\alpha}(t) = h(t)$ . Therefore,  $\|\rho_{\alpha}(u) - h(u)\|_2 \leq K \|\alpha - q\|_2 \exp(Ku)$ .

Let  $\varepsilon \in \mathbb{R}_{> 0}$  and let  $\delta_1, \delta_2 \in \mathbb{R}_{> 0}$  be such that  $K\delta_1 \exp(Ku) \leq \varepsilon$  and the open ball centered at  $q$  of radius  $\delta_2$  is contained in  $U$ . Let  $\delta = \min(\delta_1, \delta_2)$ . Since  $q$  is an  $\omega$ -limit point of  $f$ , there exists a sequence of non-negative reals  $\{t_i\}_{i \in \mathbb{Z}_{> 0}}$  such that  $t_i \rightarrow \infty$  as  $i \rightarrow \infty$  and for all  $i \in \mathbb{Z}_{> 0}$ ,  $\|f(t_i) - q\|_2 < \delta$ . Then for all  $i \in \mathbb{Z}_{> 0}$ ,  $f(t_i) \in U$ , so by Lemma 3 for all  $t \in [0, u]$ ,  $f(t_i + t) = \rho_{f(t_i)}(t)$ . Then

$$\begin{aligned} \|f(t_i + u) - h(u)\|_2 &= \|\rho_{f(t_i)}(u) - h(u)\|_2 \\ &\leq K \|f(t_i) - q\|_2 \exp(Ku) \\ &\leq K \delta \exp(Ku) \\ &\leq \varepsilon \end{aligned}$$

Thus  $h(u)$  is an  $\omega$ -limit point of  $f$ , so  $S$  is an invariant set of  $\mathcal{E}$ .

## 1.5 Finite Natural Event-Systems

In this section, we focus on finite, natural event-systems—a subclass of finite, physical event-systems which has much in common with systems of chemical reactions that obey detailed balance.

In chemical reactions, the total bond energy of the reactants minus the total bond energy of the products is a measure of the heat released. For example, in the reaction,  $\sigma X_2 - \tau X_1$ ,  $\ln\left(\frac{\sigma}{\tau}\right)$  is taken to be the quantity of heat released. If there are multiple reaction paths that take the same reactants to the same products, then the quantity of heat released along each path must be the same.

The finite, physical event-system  $\mathcal{E} = \{2X_2 - X_1, X_2 - X_1\}$  does not behave like a chemical reaction system since, when  $X_2$  is converted to  $X_1$  by the first reaction,

In (2) units of heat are released; however, when  $X_2$  is converted to  $X_1$  by the second reaction,  $\ln(1) = 0$  units of heat are released. When an event-system admits a pair of paths from the same reactants to the same products but with different quantities of heat released, we say that the system has an “energy cycle.”

**Definition 24** (*Energy cycle*) Let  $\mathcal{E}$  be a finite, physical event-system.  $\mathcal{E}$  has an energy cycle iff  $G_{\mathcal{E}}$  has a cycle of non-zero weight.

*Example 7* For the physical event-system  $\mathcal{E}_1 = \{2X_2 - X_1, X_2 - X_1\}$ , the event  $X_2 - X_1$  induces an edge  $\langle X_2, X_1 \rangle$  in the event graph with weight  $\ln\left(\frac{1}{1}\right) = 0$ . The event  $2X_2 - X_1$  induces an edge  $\langle X_1, X_2 \rangle$  with weight  $-\ln\left(\frac{2}{1}\right) = -\ln(2)$ . The weight of the cycle from  $X_2$  to  $X_1$  and back to  $X_2$  using these two edges, is  $-\ln(2) \neq 0$ . Hence,  $\mathcal{E}_1$  has an energy cycle by Definition 24.

*Example 8* For the physical event-system  $\mathcal{E}_2 = \{X_2 - X_1, 2X_3X_4 - X_2X_3, X_4X_5 - X_1X_5\}$ , the cycle  $\langle X_3X_4X_5, X_2X_3X_5, X_1X_3X_5, X_3X_4X_5 \rangle$  is induced by the sequence of events  $2X_3X_4 - X_2X_3, X_2 - X_1, X_4X_5 - X_1X_5$  and has corresponding weight  $\ln\frac{2}{1} + \ln\frac{1}{1} + \ln\frac{1}{1} = \ln(2) \neq 0$ . Hence,  $\mathcal{E}_2$  has an energy cycle.

The following theorem gives multiple characterizations of natural event-systems.

**Theorem 5** *Let  $\mathcal{E}$  be a finite, physical event-system of dimension  $n$ . The following are equivalent:*

1.  $\mathcal{E}$  is natural.
2.  $\mathcal{E}$  has a strong equilibrium point that is not a  $z$ -point. (i.e. there exists  $\alpha \in \mathbb{C}^n$  such that for all  $i = 1$  to  $n$ ,  $\alpha_i \neq 0$  and for all  $e \in \mathcal{E}$ ,  $e(\alpha) = 0$ .)
3.  $\mathcal{E}$  has no energy cycles.
4. If  $\mathcal{E} = \{\sigma_1M_1 - \tau_1N_1, \sigma_2M_2 - \tau_2N_2, \dots, \sigma_mM_m - \tau_mN_m\}$  and for all  $j = 1$  to  $m$ ,  $M_j < N_j$  and  $\sigma_j, \tau_j > 0$  then there exists  $\alpha \in \mathbb{R}^n$  such that

$$\Gamma_{\mathcal{E}}\alpha = \left\langle \ln\left(\frac{\sigma_1}{\tau_1}\right), \dots, \ln\left(\frac{\sigma_m}{\tau_m}\right) \right\rangle^T.$$

To prove Theorem 5, we will use the following lemma.

**Lemma 11** *Let  $\mathcal{E} = \{\sigma_1M_1 - \tau_1N_1, \sigma_2M_2 - \tau_2N_2, \dots, \sigma_mM_m - \tau_mN_m\}$  be a finite, physical event-system of dimension  $n$  such that for all  $j = 1$  to  $m$ ,  $\sigma_j, \tau_j > 0$  and  $M_j < N_j$ . Then for all  $\alpha = \langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle^T \in \mathbb{R}^n$ ,*

$$\Gamma_{\mathcal{E}} \cdot \alpha = \left\langle \ln\left(\frac{\sigma_1}{\tau_1}\right), \ln\left(\frac{\sigma_2}{\tau_2}\right), \dots, \ln\left(\frac{\sigma_m}{\tau_m}\right) \right\rangle^T$$

*iff  $\langle e^{\alpha_1}, \dots, e^{\alpha_n} \rangle$  is a positive strong  $\mathcal{E}$ -equilibrium point.*

*Proof* Let  $\mathcal{E} = \{\sigma_1M_1 - \tau_1N_1, \sigma_2M_2 - \tau_2N_2, \dots, \sigma_mM_m - \tau_mN_m\}$  and for all  $j = 1$  to  $m$ ,  $M_j < N_j$  and  $\sigma_j, \tau_j > 0$ . Let  $\Gamma = \Gamma_{\mathcal{E}}$ . For all  $\alpha = \langle \alpha_1, \dots, \alpha_n \rangle \in \mathbb{R}^n$ ,

$$\begin{aligned}
\Gamma\alpha &= \left\langle \ln\left(\frac{\sigma_1}{\tau_1}\right), \ln\left(\frac{\sigma_2}{\tau_2}\right), \dots, \ln\left(\frac{\sigma_m}{\tau_m}\right) \right\rangle^T \\
&\Leftrightarrow \sum_{i=1}^n \gamma_{j,i} \alpha_i = \ln(\sigma_j/\tau_j), \quad \forall j = 1, 2, \dots, m \\
&\Leftrightarrow \prod_{i=1}^n (e^{\alpha_i})^{\gamma_{j,i}} = \sigma_j/\tau_j, \quad \forall j = 1, 2, \dots, m \quad (\text{Exponentiation.}) \\
&\Leftrightarrow N_j((e^{\alpha_1}, \dots, e^{\alpha_n}) / M_j((e^{\alpha_1}, \dots, e^{\alpha_n}))) = \sigma_j/\tau_j, \quad \forall j = 1, 2, \dots, m \\
&\quad (\text{Definition of } \Gamma.) \\
&\Leftrightarrow \sigma_j M_j((e^{\alpha_1}, \dots, e^{\alpha_n})) - \tau_j N_j((e^{\alpha_1}, \dots, e^{\alpha_n})) = 0, \quad \forall j = 1, 2, \dots, m \\
&\Leftrightarrow (e^{\alpha_1}, \dots, e^{\alpha_n}) \text{ is a positive strong } \mathcal{E}\text{-equilibrium point.}
\end{aligned}$$

*Proof* (Proof of Theorem 5) (4)  $\Rightarrow$  (1) : Follows from Lemma 11.

(1)  $\Rightarrow$  (2) : Follows immediately from definitions.

(2)  $\Rightarrow$  (3) :

Consider an arbitrary cycle  $\mathcal{C}$  in  $G_{\mathcal{E}}$  given by the sequence of  $k$  edges  $\{\langle v_0, v_1 \rangle, \langle v_1, v_2 \rangle, \dots, \langle v_{k-1}, v_k = v_0 \rangle\}$  with corresponding weights  $r_1, r_2, \dots, r_k$ . By Definition 9, for  $i = 1, 2, \dots, k$ , there exist  $T_i \in \mathbb{M}_{\infty}$  and  $e_i \in \mathcal{E}$  with  $e_i = \sigma_i M_i - \tau_i N_i$  where  $\sigma_i, \tau_i > 0$  and  $M_i, N_i \in \mathbb{M}_{\infty}$  and  $M_i < N_i$  such that either

(1)  $v_{i-1} = T_i M_i$  and  $v_i = T_i N_i$  and  $r_i = \ln \frac{\sigma_i}{\tau_i} \in w(\langle v_{i-1}, v_i \rangle)$  or

(2)  $v_{i-1} = T_i N_i$  and  $v_i = T_i M_i$  and  $r_i = -\ln \frac{\sigma_i}{\tau_i} \in w(\langle v_{i-1}, v_i \rangle)$

Hence, there exists a vector  $\mathbf{b} = \langle b_1, b_2, \dots, b_k \rangle$  with  $b_i = 0$  or 1 such that:

$$\prod_{i=1}^k M_i^{b_i} N_i^{1-b_i} = \prod_{i=1}^k M_i^{1-b_i} N_i^{b_i} \quad (1.11)$$

$$w(\mathcal{C}) = \sum_{i=1}^k r_i = \sum_{i=1}^k (2b_i - 1) \ln\left(\frac{\sigma_i}{\tau_i}\right) \quad (1.12)$$

Let  $\alpha$  be a strong equilibrium point of  $\mathcal{E}$  that is not a z-point. Then, by Definition 7, for  $i = 1$  to  $k$ ,  $\sigma_i M_i(\alpha) - \tau_i N_i(\alpha) = 0$

$\Rightarrow \sigma_i M_i(\alpha) = \tau_i N_i(\alpha)$  for  $i = 1$  to  $k$

$\Rightarrow (\sigma_i M_i(\alpha))^{b_i} = (\tau_i N_i(\alpha))^{b_i}$  and  $(\tau_i N_i(\alpha))^{1-b_i} = (\sigma_i M_i(\alpha))^{1-b_i}$  for  $i = 1$  to  $k$

$\Rightarrow (\sigma_i M_i(\alpha))^{b_i} (\tau_i N_i(\alpha))^{1-b_i} = (\sigma_i M_i(\alpha))^{1-b_i} (\tau_i N_i(\alpha))^{b_i}$  for  $i = 1$  to  $k$

$\Rightarrow \prod_{i=1}^k (\sigma_i M_i(\alpha))^{b_i} (\tau_i N_i(\alpha))^{1-b_i} = \prod_{i=1}^k (\sigma_i M_i(\alpha))^{1-b_i} (\tau_i N_i(\alpha))^{b_i}$

$\Rightarrow \prod_{i=1}^k \sigma_i^{b_i} \tau_i^{1-b_i} = \prod_{i=1}^k \sigma_i^{1-b_i} \tau_i^{b_i}$  [From Eq. (1.1) and since  $\alpha$  is not a z-point]

$\Rightarrow \prod_{i=1}^k \frac{\sigma_i^{b_i} \tau_i^{1-b_i}}{\sigma_i^{1-b_i} \tau_i^{b_i}} = 1$

$\Rightarrow \sum_{i=1}^k (2b_i - 1) \ln\left(\frac{\sigma_i}{\tau_i}\right) = 0$  [Taking logarithm]

$\Rightarrow w(\mathcal{C}) = 0$  [From Eq. (1.2)]

Hence,  $\mathcal{E}$  has no energy cycle.

(3)  $\Rightarrow$  (4) :

Let  $\mathcal{E} = \{\sigma_1 M_1 - \tau_1 N_1, \sigma_2 M_2 - \tau_2 N_2, \dots, \sigma_m M_m - \tau_m N_m\}$  and for all  $j = 1$  to  $m$ ,  $M_j < N_j$  and  $\sigma_j, \tau_j > 0$ . Let  $\Gamma = \Gamma_{\mathcal{E}}$ . We shall prove that if the linear equation  $\Gamma \alpha = \langle \ln(\sigma_1/\tau_1), \dots, \ln(\sigma_m/\tau_m) \rangle^T$  has no solution in  $\mathbb{R}^n$  then  $\mathcal{E}$  has an energy cycle. For  $j = 1$  to  $m$ , let  $\Gamma_j$  be the  $j$ th row of  $\Gamma$ . If the system of linear equations  $\Gamma \alpha = \langle \ln(\sigma_1/\tau_1), \dots, \ln(\sigma_m/\tau_m) \rangle^T$  has no solution in  $\mathbb{R}^n$  then, from linear algebra [13, p. 164, Theorem] and the fact that  $\Gamma$  is a matrix of integers, it follows that there exists  $l$ , there exist (not necessarily distinct) integers  $j_1, j_2, \dots, j_l \in \{1, 2, \dots, m\}$ , there exist  $a_1, a_2, \dots, a_l \in \{+1, -1\}$  such that:

$$a_1 \Gamma_{j_1} + a_2 \Gamma_{j_2} + \dots + a_l \Gamma_{j_l} = 0 \quad (1.13)$$

$$a_1 \ln(\sigma_{j_1}/\tau_{j_1}) + a_2 \ln(\sigma_{j_2}/\tau_{j_2}) + \dots + a_l \ln(\sigma_{j_l}/\tau_{j_l}) \neq 0 \quad (1.14)$$

Consider the sequence  $\mathcal{C}$  of  $l + 1$  vertices in the event-graph defined recursively by

$$v_0 = \prod_{i=1, a_i=+1}^l M_{j_i} \prod_{i=1, a_i=-1}^l N_{j_i}$$

and for  $i = 1$  to  $l$ ,

$$v_i = \frac{v_{i-1} N_{j_i}^{a_i}}{M_{j_i}^{a_i}}$$

Observe that by (3),

$$\prod_{i=1}^l \left( \frac{N_{j_i}}{M_{j_i}} \right)^{a_i} = 1$$

Hence,

$$v_0 = \prod_{i=1, a_i=+1}^l M_{j_i}^{a_i} \prod_{i=1, a_i=-1}^l N_{j_i}^{-a_i} = \prod_{i=1, a_i=+1}^l N_{j_i}^{a_i} \prod_{i=1, a_i=-1}^l M_{j_i}^{-a_i} = v_l$$

Hence,  $\mathcal{C}$  is a cycle. Further, for  $i = 1$  to  $l$ ,

$$a_i \ln \frac{\sigma_{j_i}}{\tau_{j_i}} \in w((v_{i-1}, v_i))$$

From Eq. (1.4),

$$w(\mathcal{C}) = a_1 \ln(\sigma_{j_1}/\tau_{j_1}) + a_2 \ln(\sigma_{j_2}/\tau_{j_2}) + \dots + a_l \ln(\sigma_{j_l}/\tau_{j_l}) \neq 0$$

Hence,  $\mathcal{C}$  is an energy cycle.

Horn and Jackson [10] and Feinberg [5] have proved that chemical reaction networks with appropriate properties admit Lyapunov functions. While finite, natural event-systems are closely related to the chemical reaction networks considered by Horn and Jackson and by Feinberg, they are not identical. Consequently,

we will prove the existence of Lyapunov functions for finite, natural event-systems (Theorem 7).

The Lyapunov function is analogous in form and properties to “Entropy of the Universe” in thermodynamics. The Lyapunov function composed with an event-process is monotonic with respect to time, providing an analogy to the second law of thermodynamics.

**Definition 25** Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c = \langle c_1, c_2, \dots, c_n \rangle$ . Then  $g_{\mathcal{E},c}: \mathbb{R}_{>0}^n \rightarrow \mathbb{R}$  is given by

$$g_{\mathcal{E},c}(x_1, x_2, \dots, x_n) = \sum_{i=1}^n (x_i (\ln(x_i) - 1 - \ln(c_i)) + c_i)$$

The function  $g_{\mathcal{E},c}$  will turn out to be the desired Lyapunov function.

Note that if  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are two finite natural event-systems of the same dimension and if  $c$  is a positive strong  $\mathcal{E}_1$ -equilibrium point as well as a positive strong  $\mathcal{E}_2$ -equilibrium point, then the functions  $g_{\mathcal{E}_1,c}$  and  $g_{\mathcal{E}_2,c}$  are identical.

**Lemma 12** Let  $\mathcal{E} = \{\sigma_1 M_1 - \tau_1 N_1, \sigma_2 M_2 - \tau_2 N_2, \dots, \sigma_m M_m - \tau_m N_m\}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c$ , such that for all  $j = 1$  to  $m$ ,  $\sigma_j, \tau_j > 0$  and  $M_j < N_j$ . Then for all  $x \in \mathbb{R}_{>0}^n$ ,

$$\nabla g_{\mathcal{E},c}(x) \cdot P_{\mathcal{E}}(x) = \sum_{j=1}^m (\sigma_j M_j(x) - \tau_j N_j(x)) \ln \left( \frac{\tau_j N_j(x)}{\sigma_j M_j(x)} \right)$$

*Proof* Let  $g = g_{\mathcal{E},c}$ . Let  $x = \langle x_1, x_2, \dots, x_n \rangle \in \mathbb{R}_{>0}^n$ . Let  $P = P_{\mathcal{E}}$ .

$$\begin{aligned} \nabla g(x) \cdot P(x) &= \sum_{i=1}^n \left( \frac{\partial g}{\partial x_i}(x) \cdot P_i(x) \right) \\ &= \sum_{i=1}^n \ln \left( \frac{x_i}{c_i} \right) \left( \sum_{j=1}^m \gamma_{j,i} (\sigma_j M_j(x) - \tau_j N_j(x)) \right) \\ &= \sum_{j=1}^m (\sigma_j M_j(x) - \tau_j N_j(x)) \sum_{i=1}^n \ln \left( \left( \frac{x_i}{c_i} \right)^{\gamma_{j,i}} \right) \\ &= \sum_{j=1}^m (\sigma_j M_j(x) - \tau_j N_j(x)) \ln \left( \prod_{i=1}^n \left( \frac{x_i}{c_i} \right)^{\gamma_{j,i}} \right) \\ &= \sum_{j=1}^m (\sigma_j M_j(x) - \tau_j N_j(x)) \ln \left( \frac{\tau_j N_j(x)}{\sigma_j M_j(x)} \right) \end{aligned}$$

The last equality follows from the definition of  $\Gamma_{\mathcal{E}}$  and the fact that  $c$  is a strong-equilibrium point.

**Lemma 13** For all  $x \in \mathbb{R}_{>0}$ ,  $(1 - x) \ln(x) \leq 0$  with equality iff  $x = 1$ .

*Proof* If  $0 < x < 1$  then  $1 - x > 0$  and  $\ln(x) < 0$ . If  $x > 1$  then  $1 - x < 0$  and  $\ln(x) > 0$ . In either case, the product is strictly negative. If  $x = 1$  then  $(1 - x) \ln(x) = 0$

**Theorem 6** Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c$ . Then for all  $x \in \mathbb{R}_{>0}^n$ ,  $\nabla g_{\mathcal{E},c}(x) \cdot P_{\mathcal{E}}(x) \leq 0$  with equality iff  $x$  is a strong  $\mathcal{E}$ -equilibrium point.

*Proof* Let  $\mathcal{E} = \{\sigma_1 M_1 - \tau_1 N_1, \sigma_2 M_2 - \tau_2 N_2, \dots, \sigma_m M_m - \tau_m N_m\}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c$ , such that for all  $j = 1$  to  $m$ ,  $\sigma_j, \tau_j > 0$  and  $M_j < N_j$ . Let  $P = P_{\mathcal{E}}$  and let  $g = g_{\mathcal{E},c}$ . By Lemma 12, for all  $x \in \mathbb{R}_{>0}^n$ ,

$$\nabla g(x) \cdot P(x) = \sum_{j=1}^m (\sigma_j M_j(x) - \tau_j N_j(x)) \ln \left( \frac{\tau_j N_j(x)}{\sigma_j M_j(x)} \right)$$

From Lemma 13 and the observation that for  $j = 1, 2, \dots, m$ ,  $M_j(x), N_j(x) > 0$  when  $x \in \mathbb{R}_{>0}^n$  and by assumption  $\sigma_j, \tau_j > 0$ , we have,

$$\nabla g(x) \cdot P(x) \leq 0$$

with equality iff for all  $j = 1, 2, \dots, m$ ,  $\sigma_j M_j(x) = \tau_j N_j(x)$ . This occurs iff  $x$  is a strong  $\mathcal{E}$ -equilibrium point.

Recall that a function  $g$  is a Lyapunov function at a point  $p$  for a vector field  $V$  iff  $g$  is smooth, positive definite at  $p$  and  $L_V g$  is negative semi-definite at  $p$  [11, p. 131]. For a finite natural event-system  $\mathcal{E}$ ,  $P_{\mathcal{E}}$  induces a vector field on  $\mathbb{R}^n$ . We will show that, if  $c$  is a positive strong  $\mathcal{E}$ -equilibrium point, then  $g_{\mathcal{E},c}$  is a Lyapunov function at  $c$  for the vector field induced by  $P_{\mathcal{E}}$ .

**Theorem 7** (Existence of Lyapunov Function) Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c$ . Then  $g_{\mathcal{E},c}$  is a Lyapunov function for the vector field induced by  $P_{\mathcal{E}}$  at  $c$ .

*Proof* Let  $g = g_{\mathcal{E},c}$ . For  $i = 1, 2, \dots, n$ :

$$\frac{\partial g}{\partial x_i} = \ln \left( \frac{x_i}{c_i} \right)$$

which are all in  $\mathcal{C}^\infty$  as functions on  $\mathbb{R}_{>0}^n$ , hence  $g$  is in  $\mathcal{C}^\infty$ .

$$\frac{\partial g}{\partial x_i}(c) = \ln \left( \frac{c_i}{c_i} \right) = 0$$

establishes that  $\nabla g(c) = 0$ . For  $i = 1, 2, \dots, n$ , for  $k = 1, 2, \dots, n$ :

$$\frac{\partial^2 g}{\partial x_k \partial x_i} = \frac{\delta_{i,k}}{x_i}$$

where  $\delta_{i,k}$  is the Kronecker delta function. Hence, for all  $x \in \mathbb{R}_{>0}^n$ , the Hessian of  $g$  at  $x$  is positive definite. Therefore,  $g$  is strictly convex over  $\mathbb{R}_{>0}^n$ . Further,  $g(c) = 0$  and  $\nabla g(c) = 0$  and  $g$  is strictly convex together imply that  $g$  is positive definite at  $c$ . To establish  $g$  as a Lyapunov function, it remains to show that the directional derivative  $L_P g$  of  $g$  in the direction of the vector field induced by  $P = P_{\mathcal{E}}$  is negative semi-definite at  $c$ . This follows from Theorem 6 since for all  $x \in \mathbb{R}_{>0}^n$ ,  $L_P g(x) = \nabla g(x) \cdot P(x) \leq 0$ .

Henceforth, the function  $g_{\mathcal{E},c}$  will be called the Lyapunov function of  $\mathcal{E}$  at  $c$ . The next theorem shows that finite, natural event-systems satisfy a form of “detailed balance.”

**Theorem 8** *If  $\mathcal{E}$  is a natural, finite event-system of dimension  $n$  then all positive  $\mathcal{E}$ -equilibrium points are strong  $\mathcal{E}$ -equilibrium points.*

*Proof* Let  $P = P_{\mathcal{E}}$ . Let  $c \in \mathbb{R}_{>0}^n$  be a positive strong  $\mathcal{E}$ -equilibrium point. Let  $x$  be a positive  $\mathcal{E}$ -equilibrium point. That is,  $P(x) = 0$ . Hence,  $\nabla g_{\mathcal{E},c}(x) \cdot P_{\mathcal{E}}(x) = 0$ . By Theorem 6,  $x$  is a strong  $\mathcal{E}$ -equilibrium point.

The following lemma was proved by Feinberg [5, Proposition B.1].

**Lemma 14** *Let  $n > 0$  be an integer. Let  $U$  be a linear subspace of  $\mathbb{R}^n$ , and let  $a = \langle a_1, a_2, \dots, a_n \rangle$  and  $b$  be elements of  $\mathbb{R}_{>0}^n$ . There is a unique element  $\mu = \langle \mu_1, \mu_2, \dots, \mu_n \rangle \in U^\perp$  such that  $\langle a_1 e^{\mu_1}, a_2 e^{\mu_2}, \dots, a_n e^{\mu_n} \rangle - b$  is an element of  $U$ .*

The next theorem follows from one proved by Horn and Jackson [10, Lemma 4B]. Our proof is derived from Feinberg’s [5, Proposition 5.1].

**Theorem 9** *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $H$  be a positive conservation class of  $\mathcal{E}$ . Then  $H$  contains exactly one positive strong  $\mathcal{E}$ -equilibrium point.*

*Proof* Let  $\Gamma = \Gamma_{\mathcal{E}}$ . Let  $c^* = \langle c_1^*, c_2^*, \dots, c_n^* \rangle$  be a positive strong  $\mathcal{E}$ -equilibrium point. Let  $P \in H \cap \mathbb{R}_{>0}^n$ . For all  $c \in \mathbb{R}_{>0}^n$ ,

(1)  $c$  is a strong  $\mathcal{E}$ -equilibrium point

$$\Leftrightarrow \Gamma \langle \ln(c_1), \ln(c_2), \dots, \ln(c_n) \rangle^T = \Gamma \langle \ln(c_1^*), \ln(c_2^*), \dots, \ln(c_n^*) \rangle^T. \text{ (Lemma 11)}$$

$$\Leftrightarrow \Gamma \left\langle \ln\left(\frac{c_1}{c_1^*}\right), \ln\left(\frac{c_2}{c_2^*}\right), \dots, \ln\left(\frac{c_n}{c_n^*}\right) \right\rangle^T = 0$$

$$\Leftrightarrow \text{There exists } \mu = \langle \mu_1, \mu_2, \dots, \mu_n \rangle \in \ker \Gamma \cap \mathbb{R}^n \text{ such that } \langle \ln\left(\frac{c_1}{c_1^*}\right), \ln\left(\frac{c_2}{c_2^*}\right), \dots, \ln\left(\frac{c_n}{c_n^*}\right) \rangle^T = \mu.$$



$\Leftrightarrow$  There exists  $\mu = \langle \mu_1, \mu_2, \dots, \mu_n \rangle \in \ker \Gamma \cap \mathbb{R}^n$  such that  $c_i = c_i^* e^{\mu_i}$  for  $i = 1, 2, \dots, n$ .

(2)  $c \in H \cap \mathbb{R}^n \Leftrightarrow c - P \in (\ker \Gamma)^\perp \cap \mathbb{R}^n$ . (By Definition 17)

From (1) and (2),  $c$  is a positive strong  $\mathcal{E}$ -equilibrium point in  $H$  iff there exists  $\mu \in \ker \Gamma \cap \mathbb{R}^n$  such that  $c = \langle c_1^* e^{\mu_1}, c_2^* e^{\mu_2}, \dots, c_n^* e^{\mu_n} \rangle$  and  $\langle c_1^* e^{\mu_1}, c_2^* e^{\mu_2}, \dots, c_n^* e^{\mu_n} \rangle - P \in (\ker \Gamma)^\perp \cap \mathbb{R}^n$ . Applying Lemma 14 with  $a = c^*$ ,  $b = P$  and  $U = (\ker \Gamma)^\perp \cap \mathbb{R}^n$ , it follows that there exists a unique  $\mu$  of the desired form. Hence, there exists a unique positive strong  $\mathcal{E}$ -equilibrium point in  $H$  given by  $c = \langle c_1^* e^{\mu_1}, c_2^* e^{\mu_2}, \dots, c_n^* e^{\mu_n} \rangle$ .

To prove the main theorem of this section (Theorem 10), we will first establish several technical lemmas.

Lemma 15 shows that an event that remains zero at all times along a process can be ignored.

**Lemma 15** *Let  $\mathcal{E}$  be a finite event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be non-empty, open and simply-connected, and let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$ . Then either for all  $t \in \Omega$ ,  $f(t)$  is a strong  $\mathcal{E}$ -equilibrium point or there exist a finite event-system  $\hat{\mathcal{E}}$  of dimension  $\hat{n} \leq n$ , an  $\hat{\mathcal{E}}$ -process  $\hat{f} = \langle \hat{f}_1, \hat{f}_2, \dots, \hat{f}_{\hat{n}} \rangle$  on  $\Omega$ , and a permutation  $\pi$  on  $\{1, 2, \dots, n\}$  such that:*

1. *If  $\mathcal{E}$  is physical then  $\hat{\mathcal{E}}$  is physical.*
2. *If  $\mathcal{E}$  is natural then  $\hat{\mathcal{E}}$  is natural.*
3. *If  $c = \langle c_1, c_2, \dots, c_n \rangle$  is a positive strong  $\mathcal{E}$ -equilibrium point, then  $\hat{c} = \langle c_{\pi^{-1}(1)}, c_{\pi^{-1}(2)}, \dots, c_{\pi^{-1}(\hat{n})} \rangle$  is a positive strong  $\hat{\mathcal{E}}$ -equilibrium point.*
4. *For all  $e \in \hat{\mathcal{E}}$ , there exists  $t \in \Omega$  such that  $e(\hat{f}(t)) \neq 0$ .*
5. *If  $\hat{\mathcal{E}}$  is natural,  $I \subseteq \Omega \cap \mathbb{R}_{\geq 0}$  is connected,  $0 \in I$  and  $f(0)$  is a non-negative point then for all  $t \in I \cap \mathbb{R}_{> 0}$ ,  $\hat{f}(t)$  is a positive point.*
6. *For  $i = 1, 2, \dots, n$ , if  $\pi(i) \leq \hat{n}$  then for all  $t \in \Omega$ ,  $f_i(t) = \hat{f}_{\pi(i)}(t)$ .*
7. *For  $i = 1, 2, \dots, n$ , if  $\pi(i) > \hat{n}$  then for all  $t_1, t_2 \in \Omega$ ,  $f_i(t_1) = f_i(t_2)$ .*

*Proof* Let  $m = |\mathcal{E}|$ . Let  $\mathcal{E}_1 = \{e \in \mathcal{E} \mid \text{there exists } t \in \Omega, e(f(t)) \neq 0\}$ . If  $\mathcal{E}_1 = \emptyset$  then for all  $t \in \Omega$ ,  $e(f(t)) = 0$ , so  $f(t)$  is a strong  $\mathcal{E}$ -equilibrium point and the Lemma holds. Assume  $\mathcal{E}_1 \neq \emptyset$  and let  $\hat{m} = |\mathcal{E}_1|$ . For  $j = 1, 2, \dots, m$ , let  $\sigma_j, \tau_j \in \mathbb{R}_{> 0}$  and  $M_j = \prod_{i=1}^n X_i^{a_{j,i}}$ ,  $N_j = \prod_{i=1}^n X_i^{b_{j,i}} \in \mathbb{M}_\infty$  be such that  $M_j < N_j$  and  $\{\sigma_1 M_1 - \tau_1 N_1, \sigma_2 M_2 - \tau_2 N_2, \dots, \sigma_{\hat{m}} M_{\hat{m}} - \tau_{\hat{m}} N_{\hat{m}}\} = \mathcal{E}_1$  and  $\{\sigma_1 M_1 - \tau_1 N_1, \sigma_2 M_2 - \tau_2 N_2, \dots, \sigma_m M_m - \tau_m N_m\} = \mathcal{E}$ .

Let  $C = \{i \mid \text{there exists } j \leq \hat{m} \text{ such that either } a_{j,i} \neq 0 \text{ or } b_{j,i} \neq 0\}$ . Let  $\hat{n} = |C|$ . Let  $\pi$  be a permutation on  $\{1, 2, \dots, n\}$  such that  $\pi(C) = \{1, 2, \dots, \hat{n}\}$ .

For  $j = 1, 2, \dots, \hat{m}$ , let  $e_{\pi,j} = \sigma_j \prod_{i=1}^{\hat{n}} X_i^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} X_i^{b_{j,\pi^{-1}(i)}}$ . Let  $\hat{\mathcal{E}} = \{e_{\pi,1}, e_{\pi,2}, \dots, e_{\pi,\hat{m}}\}$ .

It follows that  $\hat{\mathcal{E}}$  is a finite event-system of dimension  $\hat{n} \leq n$ . For  $i = 1, 2, \dots, \hat{n}$ , let  $\hat{f}_i = f_{\pi^{-1}(i)}$ . Let  $\hat{f} = \langle \hat{f}_1, \hat{f}_2, \dots, \hat{f}_{\hat{n}} \rangle$ .

Let  $(\gamma_{j,i})_{m \times n} = \Gamma_{\mathcal{E}}$ . Let  $(\hat{\gamma}_{j,i})_{\hat{m} \times \hat{n}} = \Gamma_{\hat{\mathcal{E}}}$ . It follows that for  $j = 1, 2, \dots, \hat{m}$ , for  $i = 1, 2, \dots, \hat{n}$ ,

$$\hat{\gamma}_{j,i} = b_{j,\pi^{-1}(i)} - a_{j,\pi^{-1}(i)} = \gamma_{j,\pi^{-1}(i)}. \quad (1.15)$$

We claim that  $\hat{f}$  is an  $\mathcal{E}$ -process on  $\Omega$ . To see this, for  $k = 1, 2, \dots, \hat{n}$ , for all  $t \in \Omega$  :

$$\begin{aligned} \hat{f}'_k(t) &= f'_{\pi^{-1}(k)}(t) \quad [\text{Definition of } \hat{f}_k.] \\ &= \left[ \left( \sum_{j=1}^m \gamma_{j,\pi^{-1}(k)} \left( \sigma_j \prod_{i=1}^n X_i^{a_{j,i}} - \tau_j \prod_{i=1}^n X_i^{b_{j,i}} \right) \right) \circ f \right] (t) \\ &\quad [f \text{ is an } \mathcal{E}\text{-process on } \Omega.] \\ &= \left[ \left( \sum_{j=1}^{\hat{m}} \gamma_{j,\pi^{-1}(k)} \left( \sigma_j \prod_{i=1}^n X_i^{a_{j,i}} - \tau_j \prod_{i=1}^n X_i^{b_{j,i}} \right) \right) \circ f \right] (t) \\ &\quad [\text{Definition of } \mathcal{E}_1.] \\ &= \left[ \left( \sum_{j=1}^{\hat{m}} \gamma_{j,\pi^{-1}(k)} \left( \sigma_j \prod_{i \in C} X_i^{a_{j,i}} - \tau_j \prod_{i \in C} X_i^{b_{j,i}} \right) \right) \circ f \right] (t) \\ &\quad [j \leq \hat{m}, i \notin C \Rightarrow a_{j,i} = b_{j,i} = 0.] \\ &= \left[ \left( \sum_{j=1}^{\hat{m}} \gamma_{j,\pi^{-1}(k)} \left( \sigma_j \prod_{i=1}^{\hat{n}} X_{\pi^{-1}(i)}^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} X_{\pi^{-1}(i)}^{b_{j,\pi^{-1}(i)}} \right) \right) \circ f \right] (t) \\ &\quad [\pi(C) = \{1, 2, \dots, \hat{n}\}.] \\ &= \sum_{j=1}^{\hat{m}} \gamma_{j,\pi^{-1}(k)} \left( \sigma_j \prod_{i=1}^{\hat{n}} (f_{\pi^{-1}(i)}(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (f_{\pi^{-1}(i)}(t))^{b_{j,\pi^{-1}(i)}} \right) \\ &\quad [\text{By composition.}] \\ &= \sum_{j=1}^{\hat{m}} \hat{\gamma}_{j,k} \left( \sigma_j \prod_{i=1}^{\hat{n}} (f_{\pi^{-1}(i)}(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (f_{\pi^{-1}(i)}(t))^{b_{j,\pi^{-1}(i)}} \right) \\ &\quad [\text{From (15).}] \\ &= \sum_{j=1}^{\hat{m}} \hat{\gamma}_{j,k} \left( \sigma_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{b_{j,\pi^{-1}(i)}} \right) \\ &\quad [\text{Definition of } \hat{f}_i.] \\ &= \left[ \left( \sum_{j=1}^{\hat{m}} \hat{\gamma}_{j,k} e_{\pi,j} \right) \circ \hat{f} \right] (t) \quad [\text{Definition of } e_{\pi,j}.] \end{aligned}$$

This establishes the claim.

With  $\hat{\mathcal{E}}$ ,  $\hat{n}$ ,  $\hat{f}$  and  $\pi$  as described, we will now establish (1) through (6).

1. Follows from the definition of  $\hat{\mathcal{E}}$ .
2. Follows from 3.
3. Suppose  $\mathcal{E}$  is natural. Hence, there exists a positive strong  $\mathcal{E}$ -equilibrium point  $\langle c_1, c_2, \dots, c_n \rangle$ . For  $j = 1, 2, \dots, \hat{m}$  :

$$\begin{aligned}
 & e_{\pi,j}(c_{\pi^{-1}(1)}, c_{\pi^{-1}(2)}, \dots, c_{\pi^{-1}(\hat{n})}) \\
 &= \sigma_j \prod_{i=1}^{\hat{n}} c_{\pi^{-1}(i)}^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} c_{\pi^{-1}(i)}^{b_{j,\pi^{-1}(i)}} \\
 &= \sigma_j \prod_{i \in C} c_i^{a_{j,i}} - \tau_j \prod_{i \in C} c_i^{b_{j,i}} \quad [j \leq \hat{m}, i \notin C \Rightarrow a_{j,i} = b_{j,i} = 0.] \\
 &= e_j(c_1, c_2, \dots, c_n) \\
 &= 0
 \end{aligned}$$

Hence,  $\hat{c}$  is a positive strong  $\hat{\mathcal{E}}$ -equilibrium point.

4. Suppose  $j \leq \hat{m}$ . Then for all  $t \in \Omega$  :

$$\begin{aligned}
 e_{\pi,j}(\hat{f}(t)) &= \sigma_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{b_{j,\pi^{-1}(i)}} \\
 &= \sigma_j \prod_{i=1}^{\hat{n}} (f_{\pi^{-1}(i)}(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (f_{\pi^{-1}(i)}(t))^{b_{j,\pi^{-1}(i)}} \\
 &= \sigma_j \prod_{i \in C} (f_i(t))^{a_{j,i}} - \tau_j \prod_{i \in C} (f_i(t))^{b_{j,i}} \\
 &= \sigma_j \prod_{i=1}^n (f_i(t))^{a_{j,i}} - \tau_j \prod_{i=1}^n (f_i(t))^{b_{j,i}} \\
 &\quad [j \leq \hat{m}, i \notin C \Rightarrow a_{j,i} = b_{j,i} = 0.] \\
 &= \left( \left( \sigma_j \prod_{i=1}^n X_i^{a_{j,i}} - \tau_j \prod_{i=1}^n X_i^{b_{j,i}} \right) \circ f \right) (t) \\
 &= e_j(f(t))
 \end{aligned}$$

Since  $j \leq \hat{m}$ , therefore  $e_j \in \mathcal{E}_1$  and there exists  $t \in \Omega$  such that  $e_j(f(t)) \neq 0$ .

Hence, for all  $e_{\pi,j} \in \hat{\mathcal{E}}$ , there exists  $t \in \Omega$  such that  $e_{\pi,j}(\hat{f}(t)) \neq 0$ .

5. Suppose  $\hat{\mathcal{E}}$  is natural,  $I \subseteq \Omega \cap \mathbb{R}_{\geq 0}$  is connected,  $0 \in I$  and  $f(0)$  is a non-negative point. It follows that  $\hat{f}(0)$  is a non-negative point and, from Theorem 4, for all  $t \in I$ ,  $\hat{f}(t)$  is a non-negative point. Suppose, for the sake of contradiction, that there exist  $i_0 \leq \hat{n}$  and  $t_0 \in I \cap \mathbb{R}_{> 0}$  such that  $\hat{f}_{i_0}(t_0) = 0$ . From Theorem 4 again,  $\hat{f}_{i_0}(0) = 0$  and for all  $t \in I$ :  $\hat{f}_{i_0}(t) = 0$ . Since  $I$  is an interval and  $0, t_0 \in I$ ,

$I$  contains an accumulation point. Hence, since  $\hat{f}_{i_0}$  is analytic on  $\Omega$  and  $\Omega$  is connected, for all  $t \in \Omega$  :

$$\hat{f}_{i_0}(t) = 0. \quad (1.16)$$

It follows that for all  $t \in \Omega$  :

$$0 = \hat{f}'_{i_0}(t) = \sum_{j=1}^{\hat{m}} \hat{\gamma}_{j,i_0} e_{\pi,j}(\hat{f}(t)). \quad (1.17)$$

We claim that for  $j = 1, 2, \dots, \hat{m}$ , for all  $t \in \Omega$ :  $\hat{\gamma}_{j,i_0} e_{\pi,j}(\hat{f}(t)) \geq 0$ .

Case 1: Suppose  $\hat{\gamma}_{j,i_0} = 0$ . Then  $\hat{\gamma}_{j,i_0} e_{\pi,j}(\hat{f}(t)) = 0 \geq 0$ .

Case 2: Suppose  $\hat{\gamma}_{j,i_0} > 0$ . Then  $b_{j,\pi^{-1}(i_0)} > 0$ . Hence,

$$\begin{aligned} e_{\pi,j}(\hat{f}(t)) &= \sigma_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{b_{j,\pi^{-1}(i)}} \\ &= \sigma_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{a_{j,\pi^{-1}(i)}} \quad [\text{Since } b_{j,\pi^{-1}(i_0)} \\ &> 0 \text{ and from 16, } \hat{f}_{i_0}(t) = 0.] \\ &\geq 0 \quad [\hat{f}(t) \text{ is a non-negative point, by Theorem 4}] \end{aligned}$$

Hence,  $\hat{\gamma}_{j,i_0} e_{\pi,j}(\hat{f}(t)) \geq 0$ .

Case 3: Suppose  $\hat{\gamma}_{j,i_0} < 0$ . Then  $a_{j,\pi^{-1}(i_0)} > 0$ . Hence,

$$\begin{aligned} e_{\pi,j}(\hat{f}(t)) &= \sigma_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{a_{j,\pi^{-1}(i)}} - \tau_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{b_{j,\pi^{-1}(i)}} \\ &= -\tau_j \prod_{i=1}^{\hat{n}} (\hat{f}_i(t))^{b_{j,\pi^{-1}(i)}} \quad [\text{Since } a_{j,\pi^{-1}(i_0)} \\ &> 0 \text{ and from 16, } \hat{f}_{i_0}(t) = 0.] \\ &\leq 0 \quad [\hat{f}(t) \text{ is a non-negative point, by Theorem 4}] \end{aligned}$$

Hence,  $\hat{\gamma}_{j,i_0} e_{\pi,j}(\hat{f}(t)) \geq 0$ . This completes the proof of the claim.

From 1.17 and the claim, it now follows that for  $j = 1, 2, \dots, \hat{m}$ , for all  $t \in \Omega$  :

$$\hat{\gamma}_{j,i_0} e_{\pi,j}(\hat{f}(t)) = 0 \quad (1.18)$$

Since  $i_0 \leq \hat{n}$ , there exists  $j_0 \leq \hat{m}$  such that either  $a_{j_0, i_0} \neq 0$  or  $b_{j_0, i_0} \neq 0$ . If  $\hat{\gamma}_{j_0, i_0} \neq 0$  then, from 1.18,  $e_{\pi, j_0}(\hat{f}(t)) = 0$ . If  $\hat{\gamma}_{j_0, i_0} = 0$  then, since  $\hat{\gamma}_{j_0, i_0} = b_{j_0, i_0} - a_{j_0, i_0}$ , it follows that  $a_{j_0, i_0} \neq 0$  and  $b_{j_0, i_0} \neq 0$ . Hence,  $X_{i_0}$  divides  $e_{\pi, j_0}$ . From 1.16, it follows that  $e_{\pi, j_0}(\hat{f}(t)) = 0$ . Hence, irrespective of the value of  $\hat{\gamma}_{j_0, i_0}$ , for all  $t \in \Omega$ :  $e_{\pi, j_0}(\hat{f}(t)) = 0$ . Since  $e_{\pi, j_0}$  is an element of  $\hat{\mathcal{E}}$ , this leads to a contradiction with Lemma 15.4. Hence, for all  $i \leq \hat{n}$ , for all  $t \in I \cap \mathbb{R}_{>0}$ :  $\hat{f}_i(t) > 0$ .

6. Follows from the definition of  $\hat{f}$ .
7. For  $i = 1, 2, \dots, n$ , if  $\pi(i) > \hat{n}$  then  $i \notin C$ . That is, for  $j = 1, 2, \dots, m$ :  $\gamma_{j, i} = b_{j, i} - a_{j, i} = 0 - 0 = 0$ . Hence, for all  $t \in \Omega$ :  $f'_i(t) = \sum_{j=1}^m \gamma_{j, i} e_j(f(t)) = 0$ . Hence, since  $f_i$  is analytic on  $\Omega$ , and  $\Omega$  is simply-connected, for all  $t_1, t_2 \in \Omega$ :  $f_i(t_1) = f_i(t_2)$ .

We have described, for finite, natural event-systems, Lyapunov functions on the positive orthant. We next extend the definition of these Lyapunov functions to admit values at non-negative points.

**Definition 26** Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c = \langle c_1, c_2, \dots, c_n \rangle$ . For all  $v \in \mathbb{R}_{>0}$ , let  $\overline{g}_v: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$  be such that for all  $x \in \mathbb{R}_{\geq 0}$

$$\overline{g}_v(x) = \begin{cases} x(\ln(x) - 1 - \ln(v)) + v, & \text{if } x > 0; \\ v, & \text{otherwise.} \end{cases} \quad (1.19)$$

Then the *extended lyapunov function*  $\overline{g}_{\mathcal{E}, c}: \mathbb{R}_{\geq 0}^n \rightarrow \mathbb{R}$  is

$$\overline{g}_{\mathcal{E}, c}(x_1, x_2, \dots, x_n) = \sum_{i=1}^n \overline{g}_{c_i}(x_i) \quad (1.20)$$

The next lemma lists some properties of extended Lyapunov functions.

**Lemma 16** Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$  with positive strong  $\mathcal{E}$ -equilibrium point  $c = \langle c_1, c_2, \dots, c_n \rangle$ . Then:

1.  $\overline{g}_{\mathcal{E}, c}$  is continuous on  $\mathbb{R}_{\geq 0}^n$ .
2. For all  $x_1, x_2, \dots, x_n \in \mathbb{R}_{\geq 0}$ ,  $\overline{g}_{\mathcal{E}, c}(x_1, x_2, \dots, x_n) \geq 0$  with equality iff  $\langle x_1, x_2, \dots, x_n \rangle = c$ .
3. For all  $r \in \mathbb{R}_{\geq 0}$ , the set  $\{x \in \mathbb{R}_{\geq 0}^n \mid \overline{g}_{\mathcal{E}, c}(x) \leq r\}$  is bounded.
4. If  $\Omega \subseteq \mathbb{C}$  is open, simply connected and such that  $0 \in \Omega$ ,  $f = \langle f_1, f_2, \dots, f_n \rangle$  is an  $\mathcal{E}$ -process on  $\Omega$  such that  $f(0)$  is a non-negative point, and  $I \subseteq \mathbb{R}_{\geq 0} \cap \Omega$  is an interval such that  $0 \in I$  then  $(\overline{g}_{\mathcal{E}, c} \circ f)$  is monotonically non-increasing on  $I$ .

*Proof* For  $i = 1, 2, \dots, n$ , let  $\overline{g}_{c_i}(x)$  be as defined in Eq. 1.19.

1. For  $i = 1, 2, \dots, n$ ,  $\overline{g_{c_i}}$  is continuous on  $\mathbb{R}_{>0}$  and  $\lim_{x \rightarrow 0^+} \overline{g_{c_i}}(x) = c_i = \overline{g_{c_i}}(0)$ , so  $\overline{g_{c_i}}$  is continuous on  $\mathbb{R}_{\geq 0}$ . Since  $\overline{g_{\mathcal{E},c}}$  is the finite sum of continuous functions on  $\mathbb{R}_{\geq 0}$ ,  $\overline{g_{\mathcal{E},c}}$  is continuous on  $\mathbb{R}_{\geq 0}^n$ .

2. Let  $j \in \{1, 2, \dots, n\}$ . Let  $\overline{g} = \overline{g_{c_j}}$ . For all  $x \in \mathbb{R}_{>0}$ ,  $\overline{g}'(x) = \ln\left(\frac{x}{c_j}\right)$ . If  $0 < x < c_j$  then, by substitution,  $\overline{g}'(x) < 0$ . Similarly, if  $x > c_j$  then  $\overline{g}'(x) > 0$ . Hence,  $\overline{g}$  is monotonically decreasing in  $(0, c_j)$  and monotonically increasing in  $(c_j, \infty)$ . From continuity of  $\overline{g}$  in  $\mathbb{R}_{\geq 0}$ , it follows that

$$\text{For all } x \in \mathbb{R}_{\geq 0}, \overline{g}(x) \geq \overline{g}(c_j) = 0 \text{ with equality iff } x = c_j. \quad (1.21)$$

From Eqs. (1.20) and (1.21), the claim follows.

3. Observe that  $\lim_{x \rightarrow +\infty} \overline{g}(x) = +\infty$ . It follows that:

$$\text{For all } r \in \mathbb{R}_{\geq 0}, \text{ the set } \{x \in \mathbb{R}_{\geq 0} \mid \overline{g}(x) \leq r\} \text{ is bounded.} \quad (1.22)$$

If  $x_1, x_2, \dots, x_n \in \mathbb{R}_{\geq 0}$  are such that  $\overline{g_{\mathcal{E},c}}(x_1, x_2, \dots, x_n) \leq r$ , it follows from Eqs. (1.20) and (1.21) that for  $i = 1, 2, \dots, n$ :  $\overline{g_{c_i}}(x_i) \leq r$ . The claim now follows from Eq. (1.22).

4. Let  $\Omega \subseteq \mathbb{C}$  be open, simply connected, and such that  $0 \in \Omega$ ; let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$  such that  $f(0)$  is a non-negative point; and let  $I \subseteq \mathbb{R}_{\geq 0} \cap \Omega$  be an interval such that  $0 \in I$ . By Lemma 15 there exists  $\hat{n}$ ,  $\hat{\mathcal{E}}$ ,  $\hat{f}$ , and  $\pi$  satisfying 15.1–15.7. Let  $\hat{c} = \langle \hat{c}_1, \hat{c}_2, \dots, \hat{c}_{\hat{n}} \rangle = \langle c_{\pi^{-1}(1)}, c_{\pi^{-1}(2)}, \dots, c_{\pi^{-1}(\hat{n})} \rangle$ . By Lemma 15.2,  $\hat{c}$  is a positive strong equilibrium point of  $\hat{\mathcal{E}}$ . Then for all  $t \in I$ ,

$$\begin{aligned} (\overline{g_{\mathcal{E},c}} \circ f)(t) &= \sum_{i=1}^n \overline{g_{c_i}}(f_i(t)) \quad [\text{Eq. (20)}] \\ &= \sum_{i:\pi(i) \leq \hat{n}} \overline{g_{c_i}}(f_i(t)) + \sum_{i:\pi(i) > \hat{n}} \overline{g_{c_i}}(f_i(t)) \\ &= \sum_{i=1}^{\hat{n}} \overline{g_{c_{\pi^{-1}(i)}}}(f_{\pi^{-1}(i)}(t)) + \sum_{i:\pi(i) > \hat{n}} \overline{g_{c_i}}(f_i(t)) \\ &= \sum_{i=1}^{\hat{n}} \overline{g_{\hat{c}_i}}(\hat{f}_i(t)) + \sum_{i:\pi(i) > \hat{n}} \overline{g_{c_i}}(f_i(t)) \\ &\quad [\text{Definition of } \hat{c} \text{ and Lemma 15.6}] \\ &= \left( \overline{g_{\hat{\mathcal{E}},\hat{c}}} \circ \hat{f} \right)(t) + \sum_{i:\pi(i) > \hat{n}} \overline{g_{c_i}}(f_i(t)) \quad [\text{Eq. (20)}] \\ &= \left( \overline{g_{\hat{\mathcal{E}},\hat{c}}} \circ \hat{f} \right)(t) + \text{constant} \quad [\text{Lemma 15.7}] \end{aligned}$$

By Definition 26, for all  $x \in \mathbb{R}_{>0}^{\hat{n}}$ ,  $\overline{g_{\hat{\mathcal{E}},\hat{c}}}(x) = g_{\hat{\mathcal{E}},\hat{c}}(x)$ . By Lemma 15.5, for all  $t \in I \cap \mathbb{R}_{>0}$ ,  $\hat{f}(t) \in \mathbb{R}_{>0}^{\hat{n}}$ . So for all  $t \in I \cap \mathbb{R}_{>0}$ ,  $\left( \overline{g_{\hat{\mathcal{E}},\hat{c}}} \circ \hat{f} \right)(t) = \left( g_{\hat{\mathcal{E}},\hat{c}} \circ \hat{f} \right)(t)$ . Then, for all  $t \in I \cap \mathbb{R}_{>0}$ ,

$$\begin{aligned} \left( \overline{g_{\hat{\mathcal{E}},\hat{c}}} \circ \hat{f} \right)'(t) &= \left( g_{\hat{\mathcal{E}},\hat{c}} \circ \hat{f} \right)'(t) \\ &= \nabla g_{\hat{\mathcal{E}},\hat{c}}(\hat{f}(t)) \cdot \hat{f}'(t) \quad [\text{Chain rule.}] \end{aligned}$$

$$\begin{aligned}
&= \nabla g_{\hat{\mathcal{E}}, \hat{c}}(\hat{f}(t)) \cdot p_{\hat{\mathcal{E}}}(\hat{f}(t)) \quad [\text{Definition 14.}] \\
&\leq 0 \quad [\text{Theorem 6.}]
\end{aligned}$$

Therefore  $(\overline{g_{\hat{\mathcal{E}}, \hat{c}}} \circ \hat{f})$  is non-increasing on  $I \cap \mathbb{R}_{>0}$ .

By Definition 14,  $\hat{f}$  is continuous on  $I$ ; by Theorem 4,  $\hat{f}(I) \subseteq \mathbb{R}_{\geq 0}^{\hat{n}}$ ; and by Lemma 16.1,  $\overline{g_{\hat{\mathcal{E}}, \hat{c}}}$  is continuous on  $\mathbb{R}_{\geq 0}^{\hat{n}}$ ; so  $(\overline{g_{\hat{\mathcal{E}}, \hat{c}}} \circ \hat{f})$  is continuous on  $I$ . Therefore  $(\overline{g_{\hat{\mathcal{E}}, \hat{c}}} \circ \hat{f})$  is non-increasing on  $I$ . Thus  $(\overline{g_{\mathcal{E}, c}} \circ f)$  is a constant plus a monotonically non-increasing function on  $I$ , so  $(\overline{g_{\mathcal{E}, c}} \circ f)$  is monotonically non-increasing on  $I$ .

The next lemma makes use of properties of the extended Lyapunov function to show that  $\mathcal{E}$ -processes starting at non-negative points are uniformly bounded in forward real time.

**Lemma 17** *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $\alpha \in \mathbb{R}_{\geq 0}^n$ . There exists  $k \in \mathbb{R}_{>0}$  such that for all  $\Omega \subseteq \mathbb{C}$  open and simply connected and such that  $0 \in \Omega$ , for all  $\mathcal{E}$ -processes  $f = \langle f_1, f_2, \dots, f_n \rangle$  on  $\Omega$  such that  $f(0) = \alpha$ , for all intervals  $I \subseteq \Omega \cap \mathbb{R}_{\geq 0}$  such that  $0 \in I$ , for all  $t \in I$ , for  $i = 1, 2, \dots, n$ :  $f_i(t) \in \mathbb{R}$  and  $0 \leq f_i(t) < k$ .*

*Proof* Since  $\mathcal{E}$  is natural, let  $c \in \mathbb{R}_{>0}^n$  be a positive strong  $\mathcal{E}$ -equilibrium point. Let  $\overline{g} = \overline{g_{\mathcal{E}, c}}$ .

Let  $\ell = \overline{g}(\alpha)$ . Let  $S = \{x \in \mathbb{R}_{\geq 0}^n \mid \overline{g}(x) \leq \ell\}$ . By Lemma 16.3,  $S$  is bounded. Hence, let  $k$  be such that for all  $x \in S$ :  $|x|_{\infty} < k$ .

Let  $\Omega \subseteq \mathbb{C}$  be open, simply connected, and such that  $0 \in \Omega$ ; let  $f = \langle f_1, f_2, \dots, f_n \rangle$  be an  $\mathcal{E}$ -process on  $\Omega$  such that  $f(0) = \alpha$ ; and let  $I \subseteq \mathbb{R}_{\geq 0} \cap \Omega$  be an interval such that  $0 \in I$ .

From Theorem 4, for all  $t \in I$ , for  $i = 1, 2, \dots, n$ :  $f_i(t) \in \mathbb{R}$  and  $f_i(t) \geq 0$ .

Consider the function:

$$\overline{g} \circ f|_I : I \rightarrow \mathbb{R}$$

From Lemma 16.4, for all  $t \in I$ ,  $\overline{g} \circ f|_I$  is monotonically non-increasing on  $I$ . That is, for all  $t \in I$ ,

$$\overline{g}(f(t)) \leq \ell \tag{1.23}$$

It follows from Eq. 1.23 and the definition of  $S$  that  $f(I) \subseteq S$ . By the definition of  $k$ , it follows that for all  $t \in I$ , for  $i = 1, 2, \dots, n$ ,  $f_i(t) < k$ .

The next lemma shows that, because  $\mathcal{E}$ -processes starting at non-negative points are uniformly bounded in real time, they can be continued forever along forward real time.

**Lemma 18** (Existence and uniqueness of  $\mathcal{E}$ -process.) *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $\alpha \in \mathbb{R}_{\geq 0}^n$ . There exist a simply-connected open set  $\Omega \subseteq \mathbb{C}$ , an  $\mathcal{E}$ -process  $f = \langle f_1, f_2, \dots, f_n \rangle$  on  $\Omega$  and  $k \in \mathbb{R}_{>0}$  such that:*

1.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ .
2.  $f(0) = \alpha$ .
3. For all  $t \in \mathbb{R}_{\geq 0}$ , for  $i = 1, 2, \dots, n$ :  $f_i(t) \in \mathbb{R}$  and  $0 \leq f_i(t) < k$ .
4. For all simply-connected open sets  $\tilde{\Omega} \subseteq \mathbb{C}$ , for all  $\mathcal{E}$ -processes  $\tilde{f}$  on  $\tilde{\Omega}$ , for all intervals  $I \subseteq \tilde{\Omega} \cap \mathbb{R}_{\geq 0}$ , if  $0 \in I$  and  $\tilde{f}(0) = \alpha$ , then for all  $t \in I$ ,  $f(t) = \tilde{f}(t)$ .

*Proof Claim:* There exists  $k \in \mathbb{R}_{> 0}$  such that for all intervals  $I \subseteq \mathbb{R}_{\geq 0}$  with  $0 \in I$ , for all real- $\mathcal{E}$ -processes  $\tilde{h} = \langle \tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n \rangle$  on  $I$  with  $\tilde{h}(0) = \alpha$ , for all  $t \in I$ , for  $i = 1, 2, \dots, n$ :  $0 \leq \tilde{h}_i(t) \leq k$ .

To see this, let  $I \subseteq \mathbb{R}_{\geq 0}$  be an interval such that  $0 \in I$ . Let  $\tilde{h} = \langle \tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n \rangle$  be a real- $\mathcal{E}$ -process on  $I$  such that  $\tilde{h}(0) = \alpha$ .

From Lemma 4, there exist an open, simply-connected  $\tilde{\Omega} \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $\tilde{f} = \langle \tilde{f}_1, \tilde{f}_2, \dots, \tilde{f}_n \rangle$  on  $\tilde{\Omega}$  such that:

1.  $I \subset \tilde{\Omega}$
2. For all  $t \in I$ :  $\tilde{f}(t) = \tilde{h}(t)$ .

From Lemma 17, there exists  $k \in \mathbb{R}_{> 0}$  such that for all  $t \in I$ , for  $i = 1, 2, \dots, n$ :  $\tilde{f}_i(t) \in \mathbb{R}$  and  $0 \leq \tilde{f}_i(t) < k$ . That is, for all  $t \in I$ , for  $i = 1, 2, \dots, n$ :  $0 \leq \tilde{h}_i(t) < k$ . This proves the claim.

Therefore, by [8, p. 397, Corollary], there exists  $k \in \mathbb{R}_{> 0}$ , there is a real- $\mathcal{E}$ -process  $h = \langle h_1, h_2, \dots, h_n \rangle$  on  $\mathbb{R}_{\geq 0}$  such that  $h(0) = \alpha$  and for all  $t \in \mathbb{R}_{\geq 0}$ , for  $i = 1, 2, \dots, n$ :  $0 \leq h_i(t) < k$ . By Lemma 4, there exist an open, simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $f$  on  $\Omega$  such that  $\mathbb{R}_{\geq 0} \subseteq \Omega$  and for all  $t \in \mathbb{R}_{\geq 0}$ ,  $f(t) = h(t)$ . Therefore, for all  $t \in \mathbb{R}_{\geq 0}$ , for  $i = 1, 2, \dots, n$ :  $f_i(t) \in \mathbb{R}$  and  $0 \leq f_i(t) < k$ . Hence, Parts (1, 2, 3) are established. Part(4) follows from Lemma 2.

The next lemma shows that the  $\omega$ -limit points of  $\mathcal{E}$ -processes that start at non-negative points satisfy detailed balance.

**Lemma 19** *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ , let  $\Omega \subseteq \mathbb{C}$  be open and simply-connected, let  $f$  be an  $\mathcal{E}$ -process on  $\Omega$ , and let  $q \in \mathbb{C}^n$ . If  $\mathbb{R}_{\geq 0} \subseteq \Omega$  and  $f(0)$  is a non-negative point and  $q$  is an  $\omega$ -limit point of  $f$ , then  $q \in \mathbb{R}_{\geq 0}^n$  and is a strong  $\mathcal{E}$ -equilibrium point.*

*Proof* Suppose  $\mathbb{R}_{\geq 0} \subseteq \Omega$ ,  $f(0)$  is a non-negative point,  $S$  is the set of  $\omega$ -limit points of  $f$ , and  $q \in S$ . By Lemma 10  $q \in \mathbb{R}_{\geq 0}^n$ . By Lemma 18 there exists an open, simply-connected  $\Omega_q \subseteq \mathbb{C}$  such that  $\mathbb{R}_{\geq 0} \subseteq \Omega_q$  and an  $\mathcal{E}$ -process  $h = \langle h_1, h_2, \dots, h_n \rangle$  on  $\Omega_q$  such that  $h(0) = q$ .

Let  $c$  be a positive strong  $\mathcal{E}$ -equilibrium point. By Lemma 16.2,  $\overline{g_{\mathcal{E},c}}(f(t))$  is bounded below and, by Lemma 16.4, is monotonically non-increasing on  $\mathbb{R}_{\geq 0}$ . Therefore  $\lim_{t \rightarrow \infty} \overline{g_{\mathcal{E},c}}(f(t))$  exists. Since  $\overline{g_{\mathcal{E},c}}$  is continuous, for all  $\alpha \in S$ ,  $\overline{g_{\mathcal{E},c}}(\alpha) = \lim_{t \rightarrow \infty} \overline{g_{\mathcal{E},c}}(f(t))$ . By Lemma 10, for all  $t \in \mathbb{R}_{\geq 0}$ ,  $h(t) \in S$ . Hence,  $\overline{g_{\mathcal{E},c}}(h(t))$  is constant on  $\mathbb{R}_{\geq 0}$ .

By Lemma 15 either  $q$  is a strong  $\mathcal{E}$ -equilibrium or there exists a finite event-system  $\hat{\mathcal{E}}$  of dimension  $\hat{n} \leq n$ , an  $\hat{\mathcal{E}}$ -process  $\hat{h} = \langle \hat{h}_1, \hat{h}_2, \dots, \hat{h}_{\hat{n}} \rangle$  on  $\Omega_q$ , and a permutation  $\pi$  on  $\{1, 2, \dots, n\}$  satisfying 1–7 of Lemma 15.



Assume  $q$  is not a strong  $\mathcal{E}$ -equilibrium point. By Lemma 15.6, for  $i = 1, 2, \dots, \hat{n}$ , for all  $t \in \Omega_q$ ,  $\hat{h}_i(t) = h_{\pi^{-1}(i)}(t)$ . Let  $\hat{c} = \langle \hat{c}_1, \hat{c}_2, \dots, \hat{c}_{\hat{n}} \rangle = \langle c_{\pi^{-1}(1)}, c_{\pi^{-1}(2)}, \dots, c_{\pi^{-1}(\hat{n})} \rangle$ . By Lemma 15.3,  $\hat{c}$  is an  $\hat{\mathcal{E}}$ -strong equilibrium point.

For all  $v \in \mathbb{R}_{>0}$ , let  $\overline{g}_v$  be as defined in Eq. 1.19 in Definition 26. Then for all  $t \in \mathbb{R}_{\geq 0}$ ,

$$\begin{aligned} \overline{g_{\mathcal{E},c}}(h(t)) - \overline{g_{\hat{\mathcal{E}},\hat{c}}}(\hat{h}(t)) &= \sum_{i=1}^n \overline{g_{c_i}}(h_i(t)) - \sum_{j=1}^{\hat{n}} \overline{g_{\hat{c}_j}}(\hat{h}_j(t)) \\ &= \sum_{i=1}^n \overline{g_{c_i}}(h_i(t)) - \sum_{j=1}^{\hat{n}} \overline{g_{c_{\pi^{-1}(j)}}}(h_{\pi^{-1}(j)}(t)) \\ &= \sum_{i=1}^n \overline{g_{c_{\pi^{-1}(i)}}}(h_{\pi^{-1}(i)}(t)) - \sum_{j=1}^{\hat{n}} \overline{g_{c_{\pi^{-1}(j)}}}(h_{\pi^{-1}(j)}(t)) \\ &= \sum_{i=\hat{n}+1}^n \overline{g_{c_{\pi^{-1}(i)}}}(h_{\pi^{-1}(i)}(t)) \end{aligned}$$

But, by Lemma 15.7, if  $\pi(i) > \hat{n}$  then  $h_i(t)$  is constant. Hence,  $\overline{g_{c_{\pi^{-1}(i)}}}(h_{\pi^{-1}(i)}(t))$  is constant for  $i = \hat{n} + 1, \hat{n} + 2, \dots, n$ , so  $\overline{g_{\mathcal{E},c}}(h(t)) - \overline{g_{\hat{\mathcal{E}},\hat{c}}}(\hat{h}(t))$  is constant. Since  $\overline{g_{\mathcal{E},c}}(h(t))$  and  $\overline{g_{\mathcal{E},c}}(h(t)) - \overline{g_{\hat{\mathcal{E}},\hat{c}}}(\hat{h}(t))$  are both constant,  $\overline{g_{\hat{\mathcal{E}},\hat{c}}}(\hat{h}(t))$  must be constant. By Lemma 15.5, for all  $t \in \mathbb{R}_{>0}$ ,  $\hat{h}(t)$  is a positive point, so by Definitions 25 and 26,  $\overline{g_{\hat{\mathcal{E}},\hat{c}}}(\hat{h}(t)) = g_{\hat{\mathcal{E}},\hat{c}}(\hat{h}(t))$ . Since  $g_{\hat{\mathcal{E}},\hat{c}}(\hat{h}(t))$  is constant,  $\frac{d}{dt}g_{\hat{\mathcal{E}},\hat{c}}(\hat{h}(t)) = \nabla g_{\hat{\mathcal{E}},\hat{c}}(\hat{h}(t)) \cdot p_{\mathcal{E}}(\hat{h}(t)) = 0$ . Then by Theorem 6 and continuity  $\hat{h}(0)$  must be a strong  $\hat{\mathcal{E}}$ -equilibrium point, so for all  $e \in \hat{\mathcal{E}}$ , for all  $t \in \Omega_q$ ,  $e(\hat{h}(t)) = 0$ , which contradicts Lemma 15.4. Therefore  $q$  is a strong  $\mathcal{E}$ -equilibrium point.

The next theorem consolidates our results concerning natural event-systems. It also establishes that positive strong equilibrium points are locally attractive relative to their conservation classes. Together with the existence of a Lyapunov function, this implies that positive strong equilibrium points are asymptotically stable relative to their conservation classes [11, Theorem 5.57].

**Theorem 10** *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $H$  be a positive conservation class of  $\mathcal{E}$ . Then:*

1. *For all  $x \in H \cap \mathbb{R}_{\geq 0}^n$ , there exist  $k \in \mathbb{R}_{\geq 0}$ , an open, simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $f = \langle f_1, f_2, \dots, f_n \rangle$  on  $\Omega$  such that:*
  - a.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ .
  - b.  $f(0) = x$ .

- c. For all  $t \in \mathbb{R}_{\geq 0}$ ,  $f(t) \in H \cap \mathbb{R}_{\geq 0}^n$ .
- d. For all  $t \in \mathbb{R}_{\geq 0}$ , for  $i = 1, 2, \dots, n$ ,  $0 \leq f_i(t) \leq k$ .
- e. For all open, simply-connected  $\tilde{\Omega} \subseteq \mathbb{C}$ , for all  $\mathcal{E}$ -processes  $\tilde{f}$  on  $\tilde{\Omega}$ , if  $0 \in \tilde{\Omega}$  and  $\tilde{f}(0) = x$  then for all intervals  $I \subseteq \tilde{\Omega} \cap \mathbb{R}_{\geq 0}$  such that  $0 \in I$ , for all  $t \in I$ :  $f(t) = \tilde{f}(t)$ .

2. There exists  $c \in H$  such that:

- a.  $c$  is a positive strong  $\mathcal{E}$ -equilibrium point.
- b. For all  $d \in H$ , if  $d$  is a positive strong  $\mathcal{E}$ -equilibrium point, then  $d = c$ .
- a. There exists  $U \subseteq H \cap \mathbb{R}_{> 0}^n$  such that
  - i.  $U$  is open in  $H \cap \mathbb{R}_{> 0}^n$ .
  - ii.  $c \in U$ .
  - iii. For all  $x \in U$ , there exist an open, simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $f$  on  $\Omega$  such that
    - A.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ .
    - B.  $f(0) = x$ .
    - C.  $f(t) \rightarrow c$  as  $t \rightarrow \infty$  along the positive real line. (i.e. for all  $\varepsilon \in \mathbb{R}_{> 0}$ , there exists  $t_0 \in \mathbb{R}_{> 0}$  such that for all  $t \in \mathbb{R}_{> t_0}$ :  $\|f(t) - c\|_2 < \varepsilon$ .)

*Proof* 1. Follows from Lemma 18 and Theorem 3.

2a and 2b follow from Theorem 9.

2c Let  $c \in H$  be a positive strong- $\mathcal{E}$ -equilibrium point as in Theorem 10.2a. Let  $g = g_{\mathcal{E},c}$ . Let  $T = H \cap \mathbb{R}_{> 0}^n$ . For all  $x \in H \cap \mathbb{R}^n$ , for all  $r \in \mathbb{R}_{> 0}$ , let

$$\begin{aligned} B_r(x) &= \{y \in H \cap \mathbb{R}^n \mid \|x - y\|_2 < r\} \\ S_r(x) &= \{y \in H \cap \mathbb{R}^n \mid \|x - y\|_2 = r\} \\ \overline{B_r(x)} &= \{y \in H \cap \mathbb{R}^n \mid \|x - y\|_2 \leq r\} \end{aligned}$$

Since  $\mathbb{R}_{> 0}^n$  is open in  $\mathbb{R}^n$ , it follows that  $T$  is open in  $H \cap \mathbb{R}^n$ . Therefore, there exists  $\delta \in \mathbb{R}_{> 0}$  such that  $B_{2\delta}(c) \subseteq T$ . Let  $\delta \in \mathbb{R}_{> 0}$  be such that  $B_{2\delta}(c) \subseteq T$ . It follows that  $\overline{B_\delta(c)} \subseteq T$ .

Since  $g$  is continuous and  $S_\delta(c)$  is compact, let  $\mathbf{x}_0 \in S_\delta(c)$  be such that  $g(\mathbf{x}_0) = \inf_{x \in S_\delta(c)} g(x)$ . Let  $U = B_\delta(c) \cap \{x \in T \mid g(x) < g(\mathbf{x}_0)\}$ . It follows that  $U$  is open in  $T$ . Since  $\mathbf{x}_0 \neq c$ , and by Lemma 16.2,  $g(\mathbf{x}_0) = \overline{g_{\mathcal{E},c}(\mathbf{x}_0)} > 0 = g(c)$ . Hence,  $c \in U$ .

Let  $x \in U$ . From Lemma 18, there exist an open, simply-connected  $\Omega \subset \mathbb{C}$  and an  $\mathcal{E}$ -process  $f$  on  $\Omega$  such that  $\mathbb{R}_{\geq 0} \subseteq \Omega$  and  $f(0) = x$ .

We claim that for all  $t \in \mathbb{R}_{\geq 0}$ ,  $f(t) \in B_\delta(c)$ . Suppose not. Then there exists  $t_0 \in \mathbb{R}_{\geq 0}$  such that  $f(t_0) \in S_\delta(c)$ . From the definition of  $\mathbf{x}_0$ ,  $g(\mathbf{x}_0) \leq g(f(t_0))$ . Since  $f(0) = x \in U$ ,  $g(f(0)) < g(\mathbf{x}_0)$ . Hence,  $g(f(0)) < g(f(t_0))$ , contradicting Lemma 16.4.

To see that  $f(t) \rightarrow c$  as  $t \rightarrow \infty$  along the positive real line, suppose not. Then there exists  $\varepsilon \in \mathbb{R}_{> 0}$  such that  $\varepsilon < \delta$  and there exists an increasing sequence of

real numbers  $\{t_i \in \mathbb{R}_{>0}\}_{i \in \mathbb{Z}_{>0}}$  such that  $t_i \rightarrow \infty$  as  $i \rightarrow \infty$  and for all  $i$ ,  $f(t_i) \in \overline{B_\delta(c)} \setminus B_\varepsilon(c)$ . Since  $\overline{B_\delta(c)} \setminus B_\varepsilon(c)$  is compact, there exists a convergent subsequence. By Definition 22, the limit of this subsequence is an  $\omega$ -limit point  $q$  of  $f$  such that  $q \in \overline{B_\delta(c)} \setminus B_\varepsilon(c)$ . From Lemma 19,  $q$  is a strong- $\mathcal{E}$ -equilibrium point. Since  $q \in \overline{B_\delta(c)}$ ,  $q \in T$ . From Theorem 9,  $q = c$ . Hence,  $c \notin B_\varepsilon(c)$ , a contradiction.

We have established that positive strong equilibrium points are asymptotically stable relative to their conservation classes. A stronger result would be that if an  $\mathcal{E}$ -process starts at a positive point then it asymptotically tends to the positive strong equilibrium point in its conservation class. Such a result is related to the widely-held notion that, for systems of chemical reactions, concentrations approach equilibrium. We have been unable to prove this result. We will now state it as an open problem. This problem has a long history. It appears to have been first suggested in [10, Lemma 4C], where it was accompanied by an incorrect proof. The proof was retracted in [9].

**Open Problem 1** *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $H$  be a positive conservation class of  $\mathcal{E}$ . Then*

1. *For all  $x \in H \cap \mathbb{R}_{\geq 0}^n$ , there exist  $k \in \mathbb{R}_{\geq 0}$ , an open, simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $\tilde{f} = \langle f_1, f_2, \dots, f_n \rangle$  on  $\Omega$  such that:*
  - a.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ .
  - b.  $\tilde{f}(0) = x$ .
  - c. *For all  $t \in \mathbb{R}_{\geq 0}$ ,  $\tilde{f}(t) \in H \cap \mathbb{R}_{\geq 0}^n$ .*
  - d. *For all  $t \in \mathbb{R}_{\geq 0}$ , for  $i = 1, 2, \dots, n$ ,  $0 \leq f_i(t) < k$ .*
  - e. *For all open, simply-connected  $\tilde{\Omega} \subseteq \mathbb{C}$ , for all  $\mathcal{E}$ -processes  $\tilde{f}$  on  $\tilde{\Omega}$ , if  $0 \in \tilde{\Omega}$  and  $\tilde{f}(0) = x$  then for all intervals  $I \subseteq \tilde{\Omega} \cap \mathbb{R}_{\geq 0}$ , if  $0 \in I$  then for all  $t \in I$ :  $\tilde{f}(t) = \tilde{f}(0)$ .*
2. *There exists  $c \in H$  such that:*
  - a.  *$c$  is a positive strong  $\mathcal{E}$ -equilibrium point.*
  - b. *For all  $d \in H$ , if  $d$  is a positive strong  $\mathcal{E}$ -equilibrium point, then  $d = c$ .*
  - c. *For all  $x \in H \cap \mathbb{R}_{>0}^n$ , there exist an open, simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $f$  on  $\Omega$  such that:*
    - i.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ .
    - ii.  $f(0) = x$ .
    - iii.  *$f(t) \rightarrow c$  as  $t \rightarrow \infty$  along the positive real line. (i.e. for all  $\varepsilon \in \mathbb{R}_{>0}$ , there exists  $t_0 \in \mathbb{R}_{>0}$  such that for all  $t \in \mathbb{R}_{>t_0}$ :  $\|f(t) - c\|_2 < \varepsilon$ .)*

In light of Theorem 10, Open Problem 1 is equivalent to the following statement.

**Open Problem 2** *Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $x \in \mathbb{R}_{>0}^n$ . Then there exists an open, simply-connected  $\Omega \subseteq \mathbb{C}$ , an  $\mathcal{E}$ -process  $f$  on  $\Omega$  and a positive strong  $\mathcal{E}$ -equilibrium point  $c$  such that:*

1.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ .
2.  $f(0) = x$ .
3.  *$f(t) \rightarrow c$  as  $t \rightarrow \infty$  along the positive real line. (i.e. for all  $\varepsilon \in \mathbb{R}_{>0}$ , there exists  $t_0 \in \mathbb{R}_{>0}$  such that for all  $t \in \mathbb{R}_{>t_0}$ :  $\|f(t) - c\|_2 < \varepsilon$ .)*

## 1.6 Finite Natural Atomic Event-Systems

In this section, we settle Open 1 in the affirmative for the case of finite, natural, atomic event-systems. The atomic hypothesis appears to be a natural assumption to make concerning systems of chemical reactions. Therefore, our result may be considered a validation of the notion in chemistry that concentrations tend to equilibrium. We will prove the following theorem:

**Theorem 11** *Let  $\mathcal{E}$  be a finite, natural, atomic event-system of dimension  $n$ . Let  $\alpha \in \mathbb{R}_{>0}^n$ . Then there exists an open, simply-connected  $\Omega \subseteq \mathbb{C}$ , an  $\mathcal{E}$ -process  $f$  on  $\Omega$ , and a positive strong  $\mathcal{E}$ -equilibrium point  $c$  such that:*

1.  $\mathbb{R}_{\geq 0} \subseteq \Omega$ ,
2.  $f(0) = \alpha$ , and
3.  $f(t) \rightarrow c$  as  $t \rightarrow \infty$  along the positive real line (i.e. for all  $\varepsilon \in \mathbb{R}_{>0}$ , there exists  $t_0 \in \mathbb{R}_{>0}$  such that for all  $t \in \mathbb{R}_{>t_0}$ :  $\|f(t) - c\|_2 < \varepsilon$ ).

It follows from Theorem 10 that the point  $c$  depends only on the conservation class of  $\alpha$  and not on  $\alpha$  itself. That is, two  $\mathcal{E}$ -processes starting at positive points in the same conservation class asymptotically converge to the same  $c$ .

Implicit in the atomic hypothesis is the idea that atoms are neither created nor destroyed, but rather are conserved by chemical reactions. Our proof uses a formal analog of this idea. Recall from Definition 11 that if  $\mathcal{E}$  is atomic then  $C_{\mathcal{E}}(M)$  contains a unique monomial from  $\mathbb{M}_{A_{\mathcal{E}}}$ .

**Definition 27** *Let  $\mathcal{E}$  be a finite, natural, atomic event-system of dimension  $n$ . The atomic decomposition map  $D_{\mathcal{E}}: \mathbb{M}_{\{X_1, X_2, \dots, X_n\}} \rightarrow \mathbb{Z}_{\geq 0}^n$  is the function  $M \mapsto \langle b_1, b_2, \dots, b_n \rangle$  such that  $X_1^{b_1} X_2^{b_2} \dots X_n^{b_n} \in C_{\mathcal{E}}(M) \cap \mathbb{M}_{A_{\mathcal{E}}}$ .*

The next lemma lists some properties of the atomic decomposition map. Note that though the event-graph  $G_{\mathcal{E}}$  is directed, if  $M$  and  $N$  are monomials and there exists a path in  $G_{\mathcal{E}}$  from  $M$  to  $N$  then there also exists a path in  $G_{\mathcal{E}}$  from  $N$  to  $M$ . Informally, this is because all events are “reversible.”

**Lemma 20** *Let  $\mathcal{E}$  be a finite, natural, atomic event-system of dimension  $n$  and let  $M, N \in \mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$ . Then:*

1.  $D_{\mathcal{E}}(M) = D_{\mathcal{E}}(N)$  if and only if  $C_{\mathcal{E}}(M) = C_{\mathcal{E}}(N)$ .
2.  $D_{\mathcal{E}}(MN) = D_{\mathcal{E}}(M) + D_{\mathcal{E}}(N)$ .

*Proof* Let  $D = D_{\mathcal{E}}$ .

(1)  $D(M) = D(N) = \langle b_1, b_2, \dots, b_n \rangle$  if and only if  $X_1^{b_1} X_2^{b_2} \dots X_n^{b_n} \in C_{\mathcal{E}}(M)$  and  $X_1^{b_1} X_2^{b_2} \dots X_n^{b_n} \in C_{\mathcal{E}}(N)$ . Then  $C_{\mathcal{E}}(M) = C_{\mathcal{E}}(N)$ .

(2) Let  $D(M) = \langle b_1, b_2, \dots, b_n \rangle$  and  $D(N) = \langle c_1, c_2, \dots, c_n \rangle$ . Then, in  $G_{\mathcal{E}}$  there is a path from  $M$  to  $X_1^{b_1} X_2^{b_2} \dots X_n^{b_n} \in \mathbb{M}_{A_{\mathcal{E}}}$  and a path from  $N$  to  $X_1^{c_1} X_2^{c_2} \dots X_n^{c_n} \in \mathbb{M}_{A_{\mathcal{E}}}$ . It follows that there is a path from  $MN$  to  $X_1^{b_1+c_1} X_2^{b_2+c_2} \dots X_n^{b_n+c_n} \in \mathbb{M}_{A_{\mathcal{E}}}$ . Hence  $D(MN) = \langle b_1+c_1, b_2+c_2, \dots, b_n+c_n \rangle = D(M) + D(N)$ .

**Definition 28** Let  $\mathcal{E}$  be a finite, natural, atomic event-system of dimension  $n$ . For all  $i \in \{1, 2, \dots, n\}$ , for all  $M \in \mathbb{M}_{\{X_1, X_2, \dots, X_n\}}$ ,  $D_{\mathcal{E},i}(M)$  is the  $i$ th component of  $D_{\mathcal{E}}(M)$ .

**Definition 29** Let  $\mathcal{E}$  be a finite, natural, atomic event-system of dimension  $n$ . For all  $i \in \{1, 2, \dots, n\}$  the function  $\kappa_{\mathcal{E},i}: \mathbb{C}^n \rightarrow \mathbb{C}$  is given by

$$\langle z_1, z_2, \dots, z_n \rangle \mapsto \sum_{j=1}^n D_{\mathcal{E},i}(X_j) z_j.$$

**Lemma 21** Let  $\mathcal{E}$  be a finite, natural, atomic event-system of dimension  $n$ . Then for all  $i \in \{1, 2, \dots, n\}$ , the function  $\kappa_{\mathcal{E},i}$  is a conservation law of  $\mathcal{E}$ .

*Proof* Let  $m = |\mathcal{E}|$ , and for  $j = 1, 2, \dots, m$ , let  $\sigma_j, \tau_j \in \mathbb{R}_{>0}$  and  $M_j, N_j \in \mathbb{M}_{\infty}$  with  $M_j < N_j$  be such that  $\mathcal{E} = \{\sigma_1 M_1 - \tau_1 N_1, \dots, \sigma_m M_m - \tau_m N_m\}$ . For  $i = 1, 2, \dots, n$ , let  $a_{j,i}, b_{j,i} \in \mathbb{Z}_{>0}$  be such that  $M_j = X_1^{a_{j,1}} X_2^{a_{j,2}} \dots X_n^{a_{j,n}}$  and  $N_j = X_1^{b_{j,1}} X_2^{b_{j,2}} \dots X_n^{b_{j,n}}$ . Let  $(\gamma_{j,i})_{m \times n} = \Gamma_{\mathcal{E}}$ .

Then for  $j = 1, 2, \dots, m$ :

$$\begin{aligned} & \sigma_j M_j - \tau_j N_j \in \mathcal{E} \\ \Rightarrow & M_j \in C_{\mathcal{E}}(N_j) \quad [\text{Definition 9}] \\ \Rightarrow & D_{\mathcal{E}}(M_j) = D_{\mathcal{E}}(N_j) \quad [\text{Lemma 20}] \\ \Rightarrow & \sum_{i=1}^n a_{j,i} D_{\mathcal{E}}(X_i) = \sum_{i=1}^n b_{j,i} D_{\mathcal{E}}(X_i) \quad [\text{Lemma 20}] \\ \Rightarrow & \sum_{i=1}^n (b_{j,i} - a_{j,i}) D_{\mathcal{E}}(X_i) = 0 \\ \Rightarrow & \sum_{i=1}^n \gamma_{j,i} D_{\mathcal{E}}(X_i) = 0 \quad [\text{Definition 12}] \end{aligned}$$

It follows that for all  $j \in \{1, 2, \dots, m\}$ , for all  $k \in \{1, 2, \dots, n\}$ ,

$$\sum_{i=1}^n \gamma_{j,i} D_{\mathcal{E},k}(X_i) = 0$$

Therefore, for all  $k \in \{1, 2, \dots, n\}$ ,  $\Gamma_{\mathcal{E}} \cdot \langle D_{\mathcal{E},k}(X_1), D_{\mathcal{E},k}(X_2), \dots, D_{\mathcal{E},k}(X_n) \rangle^T = 0$ . Since the vector  $\langle D_{\mathcal{E},k}(X_1), D_{\mathcal{E},k}(X_2), \dots, D_{\mathcal{E},k}(X_n) \rangle^T$  is in the kernel of  $\Gamma_{\mathcal{E}}$ , by Theorem 2,  $\kappa_{\mathcal{E},k}$  is a conservation law of  $\mathcal{E}$ .

**Lemma 22** Let  $\mathcal{E}$  be a finite, natural event-system of dimension  $n$ . Let  $M, N \in \mathbb{M}_{\infty}$  and let  $q \in \mathbb{C}^n$ . If  $M \in C_{\mathcal{E}}(N)$  and  $q$  is a strong  $\mathcal{E}$ -equilibrium point and  $M(q) = 0$ , then  $N(q) = 0$ .

*Proof* Let  $\langle v_0, v_1 \rangle$  be an edge in  $G_{\mathcal{E}}$ . Then there exist  $e \in \mathcal{E}$  and  $\sigma, \tau \in \mathbb{R}_{>0}$  and  $T, U, V \in \mathbb{M}_{\infty}$  such that  $e = \sigma U - \tau V$  and  $v_0 = TU$  and  $v_1 = TV$ .

Assume  $v_0(q) = 0$ . Then either  $T(q) = 0$  or  $U(q) = 0$ . If  $T(q) = 0$  then  $v_1(q) = 0$ . If  $U(q) = 0$  and  $q$  is a strong  $\mathcal{E}$ -equilibrium point, then  $e(q) = \sigma U(q) - \tau V(q) = 0$ , so  $V(q) = 0$ . Therefore  $v_1(q) = 0$ . The lemma follows by induction.

We are now ready to prove Theorem 11.

*Proof* (Proof of Theorem 11) Since  $\alpha$  is a positive point, it is in some positive conservation class  $H$ . By Theorem 10:

1. There exists exactly one positive strong  $\mathcal{E}$ -equilibrium point  $c \in H$ .
2. There exist an open and simply-connected  $\Omega \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $f$  on  $\Omega$  such that  $\mathbb{R}_{\geq 0} \subset \Omega$  and  $f(0) = \alpha$ .
3. For all  $t \in \mathbb{R}_{\geq 0}$ ,  $f(t) \in H \cap \mathbb{R}_{\geq 0}^n$ .
4. There exists  $k \in \mathbb{R}_{> 0}$  such that for  $i = 1, 2, \dots, n$ , for all  $t \in \mathbb{R}_{\geq 0}$ ,  $f_i(t) \in \mathbb{R}$  and  $0 \leq f_i(t) \leq k$ .

Let  $\{t_j\}_{j \in \mathbb{Z}_{> 0}}$  be an infinite sequence of non-negative reals such that  $t_j \rightarrow \infty$  as  $j \rightarrow \infty$ . Then  $\{f(t_j)\}_{j \in \mathbb{Z}_{> 0}}$  is an infinite sequence contained in a compact subset of  $\mathbb{R}^n$ , so it must have a convergent subsequence. Let  $q = \langle q_1, q_2, \dots, q_n \rangle \in \mathbb{C}^n$  be the limit point of a convergent subsequence of  $\{f(t_j)\}_{j \in \mathbb{Z}_{> 0}}$ .  $H$  and  $\mathbb{R}_{\geq 0}^n$  are both closed in  $\mathbb{C}^n$ , so  $q \in H \cap \mathbb{R}_{\geq 0}^n$ . Since  $\mathcal{E}$  is natural and  $q$  is an  $\omega$ -limit of  $f$ ,  $q$  must be a strong  $\mathcal{E}$ -equilibrium point by Lemma 19.

Assume, for the sake of contradiction, that  $q \notin \mathbb{R}_{> 0}^n$ . Let  $i \in \{1, 2, \dots, n\}$  be such that  $q_i = 0$ . Let  $N \in C_{\mathcal{E}}(X_i) \cap \mathbb{M}_{A_{\mathcal{E}}}$ . Since  $\mathcal{E}$  is atomic, a unique such  $N$  exists. It follows from the definition of event graph that  $X_i \in C_{\mathcal{E}}(N)$ . By Lemma 22,  $N(q) = X_i(q) = q_i = 0$ . It follows that  $N \neq 1$ . Hence, there exists  $X_a \in A_{\mathcal{E}}$  such that  $X_a$  divides  $N$  and  $X_a(q) = 0$ .

For all  $j \in \{1, 2, \dots, n\}$  such that  $D_{\mathcal{E},a}(X_j) \neq 0$ , let  $M_j \in C_{\mathcal{E}}(X_j) \cap \mathbb{M}_{A_{\mathcal{E}}}$ . Then  $X_a$  divides  $M_j$ , so  $M_j(q) = 0$ . Again by Lemma 22,  $X_j(q) = M_j(q) = 0$ , so  $q_j = 0$ . It follows that for all  $j \in \{1, 2, \dots, n\}$  either  $D_{\mathcal{E},a}(X_j) = 0$  or  $q_j = 0$  so

$$\kappa_{\mathcal{E},a}(q) = \sum_{j=1}^n D_{\mathcal{E},a}(X_j)q_j = 0.$$

Since  $\kappa_{\mathcal{E},a}$  is a conservation law of  $\mathcal{E}$  by Lemma 21 and  $q$  is an  $\omega$ -limit point of  $f$ , it follows that

$$\kappa_{\mathcal{E},a}(\alpha) = 0. \quad (1.24)$$

For all  $j$ ,  $D_{\mathcal{E},a}(X_j)$  is nonnegative, and  $\alpha$  is a positive point, so for all  $j \in \{1, 2, \dots, n\}$ ,  $D_{\mathcal{E},a}(X_j)\alpha_j \geq 0$ . But  $D_{\mathcal{E},a}(X_a) = 1$  and  $\alpha_a > 0$  so  $\kappa_{\mathcal{E},a}(\alpha) > 0$ , contradicting Eq. (1.24). Therefore  $q \in \mathbb{R}_{> 0}^n$ . Since  $c$  is the unique positive strong  $\mathcal{E}$ -equilibrium point in  $H$ ,  $c = q$ .

Let  $U \subseteq H \cap \mathbb{R}_{> 0}^n$  be the open set stated to exist in Theorem 10.2c. Since  $c$  is an  $\omega$ -limit point of  $f$ , there exists  $t_0 \in \mathbb{R}_{> 0}$  such that  $f(t_0) \in U$ . Again by Theorem 10, there exist  $\tilde{\Omega} \subseteq \mathbb{C}$  and an  $\mathcal{E}$ -process  $\tilde{f}$  on  $\tilde{\Omega}$  such that  $\mathbb{R}_{\geq 0} \subseteq \tilde{\Omega}$  and  $\tilde{f}(0) = f(t_0)$  and  $\tilde{f}(t) \rightarrow c$  as  $t \rightarrow \infty$ . By Lemma 3, for all  $t \in \mathbb{R}_{\geq 0}$ ,  $f(t+t_0) = \tilde{f}(t)$ . Therefore,  $f(t) \rightarrow c$  as  $t \rightarrow \infty$ .

## 1.7 Lessons Learnt

We have endeavored to place the kinetic theory of chemical reactions on a firm mathematical foundation and to make the law of mass action available for purely mathematical consideration.

With regard to chemistry, we have proven that many of the expectations acquired through empirical study are warranted. In particular:

1. For finite event-systems, the stoichiometric coefficients determine conservation laws that processes must obey (Theorem 3). In fact, we can show
  - a. For finite, physical event-systems, the stoichiometric coefficients determine all linear conservation laws;
  - b. For finite, natural event-systems, the stoichiometric coefficients determine *all* conservation laws.
2. For finite, physical event-systems, a process begun with positive (non-negative) concentrations will retain positive (non-negative) concentrations through forward real time where it is defined (Theorem 4). For finite, natural event-systems, a process begun with positive (non-negative) concentrations will retain positive (non-negative) concentrations through *all* forward real time (Theorem 10)—that is, it will be defined through all forward real time.
3. Finite, natural event-systems must obey the “second law of thermodynamics” (Theorem 7). In addition, the flow of energy is very restrictive—finite, natural event-systems can contain no energy cycles (Theorem 5).
4. For finite, natural event-systems, every positive conservation class contains exactly one positive equilibrium point. This point is a strong equilibrium point and is asymptotically stable relative to its conservation class (Theorem 10).

Unfortunately, we, like our predecessors, are unable to settle the problem of whether a process begun with positive concentrations must approach equilibrium. We consider this the fundamental open problem in the field (Open Problem 1). For finite, natural event-systems that obey a mathematical analogue of the atomic hypothesis, we settle Open Problem 1 in the affirmative (Theorem 11). In particular, we show that for finite, natural, atomic event-systems, every positive conservation class contains exactly one non-negative equilibrium point. This point is a positive strong equilibrium point and is globally stable relative to the intersection of its conservation class with the positive orthant.

In terms of expanding the mathematical aspects of our theory, there are several potentially fruitful avenues including:

1. **Complex-analytic aspects of event-systems.** While we exploit some of the complex-analytic properties of processes in this chapter, we believe that a deeper investigation along these lines is warranted. For example, if we do not restrict the domain of a process to be simply-connected, then each component of a process becomes a complete analytic function in the sense of Weierstrass.

2. **Infinite event-systems.** Issues of convergence arise when considering infinite event-systems. To obtain a satisfactory theory, some constraints may be necessary. For example, a bound on the maximum degree of events may be worth considering. It may also be possible to generalize the notion of an atomic event-system to the infinite-dimensional case in such a way that each atom has an associated conservation law. One might then restrict initial concentrations to those for which each conservation law has a finite value. Additional constraints are likely to be needed as well.
3. **Algebraic-geometric aspects of event-systems.** Every finite event-system that generates a prime ideal has a corresponding affine toric variety (as defined in [4, p.15]). The closed points of this variety are the strong equilibria of the event-system. Further, every affine toric variety is isomorphic to an affine toric variety whose ideal is generated by a finite event system. One could generalize event-systems to allow irreversible reactions. In that case, it appears that the prime ideals generated by such event-systems are exactly the ideals corresponding to affine toric varieties.

We can show (proof not provided) that finite, natural, atomic event-systems generate prime ideals. We are working towards settling Open Problem 1 in the affirmative for every finite, natural event-system that generates a prime ideal.

**Acknowledgments** This work benefitted from discussions with many people, named here in alphabetical order: Yuliy Baryshnikov, Yuriy Brun, Qi Cheng, Ed Coffman, Ashish Goel, Jack Hale, Lila Kari, David Kempe, Eric Klavins, John Reif, Paul Rothemund, Robert Sacker, Rolfe Schmidt, Bilal Shaw, David Soloveichik, Hal Wasserman, Erik Winfree.

## References

1. Adleman L (1999) Toward a mathematical theory of self-assembly. Technical Report 00-722, University of Southern California. Department of Computer Science
2. Ahlfors L (1979) Complex analysis. McGraw-Hill, International Series in Pure and Applied Mathematics
3. Bernstein DS, Bhat SP (1999). Nonnegativity, reducibility, and semistability of mass action kinetics. In: IEEE conference on decision and control. IEEE Publications, pp 2206–2211
4. Eisenbud D, Sturmfels B (1996) Binomial ideals. *Duke Math J* 84(1):1–45
5. Feinberg M (1995) The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch Ration Mech Anal* 132:311–370
6. Gatermann K, Huber B (2002) A family of sparse polynomial systems arising in chemical reaction systems. *J Symbolic Comput* 33(3):275–305
7. Guldberg CM, Waage P (1986) Studies concerning affinity. *J Chem Educ* 63:1044
8. Hirsch MW, Stephen Smale RLD (2004) Differential equations, dynamical systems, and an introduction to chaos, 2nd edn. Elsevier Academic Press, Amsterdam
9. Horn FJM (1974) The dynamics of open reaction systems. In: Mathematical aspects of chemical and biochemical problems and quantum chemistry, vol VIII. In: Proceedings SIAM-AMS symposium on applied mathematics, New York
10. Horn FJM, Jackson R (1972) General mass action kinetics. *Arch Ration Mech Anal* 49:81–116
11. Irwin MC (1980) Smooth dynamical systems. Academic Press



12. Ito K (ed) (1987) Encyclopedic dictionary of mathematics. MIT Press, Cambridge, Massachusetts
13. Narayan S, Mittal PK (2003) A textbook of matrices, 10th edn. S. Chand and Company Ltd, New Delhi
14. Sontag ED (2001) Structure and stability of certain chemical networks and applications to the kinetic proofreading model of T-cell receptor signal transduction. *IEEE Trans Autom Control* 46:1028–1047

# Chapter 2

## Structural Analysis of Biological Networks

Franco Blanchini and Elisa Franco

**Abstract** We introduce the idea of structural analysis of biological network models. In general, mathematical representations of molecular systems are affected by parametric uncertainty: experimental validation of models is always affected by errors and intrinsic variability of biological samples. Using uncertain models for predictions is a delicate task. However, given a plausible representation of a system, it is often possible to reach general analytical conclusions on the system's admissible dynamic behaviors, regardless of specific parameter values: in other words, we say that certain behaviors are structural for a given model. Here we describe a parameter-free, qualitative modeling framework and we focus on several case studies, showing how many paradigmatic behaviors such as multistationarity or oscillations can have a structural nature. We highlight that classical control theory methods are extremely helpful in investigating structural properties.

**Keywords** Biological network · Control theory · Structural analysis · Structural property · Enzymatic networks · Jacobian · Eigenvalue · Chemical reaction network · Robustness · Set invariance · Mitogen activated protein kinase (MAPK)

### 2.1 Introduction

Structural analysis of a dynamical system aims at revealing behavioral patterns that occur regardless of the adopted parameters, or, at least, for wide parameters ranges. Due to their parametric variability, biological models are often subject to structural

---

F. Blanchini

Dipartimento di Matematica ed Informatica, Università degli Studi di Udine,  
Via delle Scienze 206, 33100 Udine, Italy  
e-mail: blanchini@uniud.it

E. Franco (✉)

Department of Mechanical Engineering, University of California at Riverside,  
900 University Avenue, Riverside, CA 92521, USA  
e-mail: efranco@engr.ucr.edu

analysis, which can be a very useful tool to reveal or rule out potential dynamic behaviors.

Even for very simple networks, simulations are the most common approach to structural investigation. For instance, three-node enzymatic networks are considered in [1], where numerical analysis shows that adaptability is mostly determined by interconnection topology rather than specific reaction parameters. In [2], through numerical exploration of the Jacobian eigenvalues for two, three and four node gene networks, the authors isolate a series of interconnections which are stable, robustly with respect to the specific parameters; the isolated structures also turn out to be the most frequent topologies in existing biological networks databases. For other examples of numerical robustness analysis, see, for instance [3–8].

Analytical approaches to the study of robustness have been proposed in specific contexts. A series of recent papers [9, 10] focused on input/output robustness of ODE models for phosphorylation cascades. In particular, the theory of chemical reaction networks is used in [10] as a powerful tool to demonstrate the property of absolute concentration robustness. Indeed, the so-called deficiency theorems are to date some of the most general results to establish robust stability of a chemical reaction network [11]. Monotonicity is also a structural property, often useful to demonstrate certain dynamic behaviors in biological models by imposing general interaction conditions [12, 13]. Robustness has also been investigated in the context of compartmental models, common in biology and biochemistry [14]. A survey on the problem of structural stability is proposed in [15].

Here we review and expand on the framework we proposed in [16], where we suggest a variety of tools for investigation of robust stability, including Lyapunov and setinvariance methods, and conditions on the network graph. We will assume that certain standard properties or assumptions are verified by our model, for example positivity, monotonicity of key interactions, and boundedness. Based on such general assumptions, we will show how dynamic behaviors can be structurally proved or ruled out for a range of examples. Our approach does not require numerical simulation efforts, and we believe that our techniques are instrumental for biological robustness analysis [17, 18].

The chapter begins with a motivating example, and a brief summary of the analysis framework in [16]. Then we consider a certain number of “paradigmatic behaviors” encountered in biochemical systems, including multistationarity, oscillations, and adaptation; through simple examples, we show how these behaviors can be deduced analytically without resorting to simulation. As relevant case studies, we consider a simplified model of the MAPK pathway and the *lac* Operon. Finally, we prove some general results on structural stability and boundedness for qualitative models that satisfy certain graphical conditions.

### 2.1.1 Motivating Example: A Qualitative Model for Transcriptional Repression

Consider a molecular system where a protein,  $x_1$ , is translated at a certain steady rate and represses the production of an RNA species  $x_2$ . In turn,  $x_2$  is the binding target of another RNA species  $u_2$  ( $x_2$  and  $u_2$  bind and form an inactive complex to be degraded); unbound  $x_2$  is translated into protein  $x_3$ . A standard parametric model is, for example, in Eq. (2.1) [19].

$$\begin{aligned}\dot{x}_1 &= k_1 u_1 - k_2 x_1, \\ \dot{x}_2 &= k_3 \frac{1}{K_1^n + x_1^n} - k_4 x_2 - k_5 x_2 u_2, \\ \dot{x}_3 &= k_6 x_2 - k_7 x_3.\end{aligned}\tag{2.1}$$

One might ask what kind of dynamic behaviors can be expected by this system. Since we cannot analytically solve these ODEs, numerical simulations would provide us with answers that depend on the parameters we believe are the most accurate in representing the physical system. Parameters might have been derived by fitting noisy data, so they are uncertain in practically all cases. The purpose of this chapter, is to highlight how we can achieve important conclusions on the potential dynamic behavior of a molecular system *without knowing the value of each parameter*.

In this specific example, we know that the system parameters are positive and bounded scalars. The Hill function  $H(x_1) = k_3 / (K_1^n + x_1^n)$  is a decreasing function, sufficiently “flat” near the origin (i.e. with zero derivative), with a single flexus (second derivative has a single zero) [19, 20]. Then, we can say that for given  $u_1$  and  $u_2$  constant or varying on a slower timescale than this system,  $x_1$  will converge to its equilibrium  $\bar{x}_1 = k_1 u_1 / k_2$ . Similarly,  $\bar{x}_2 = H(\bar{x}_1) / (k_4 + k_5 u_2)$ ,  $\bar{x}_3 = k_6 \bar{x}_2 / k_7$ . Regardless of the specific parameter values, and therefore robustly, the system is stable. While the equilibrium value for the protein  $\bar{x}_1$  could grow unbounded with  $u_1$ , the RNA species  $\bar{x}_2$  is always bounded.

## 2.2 Qualitative Models for Biological Dynamical Systems

The interactions of RNA species, proteins and biochemical ligands are at the basis of cellular development, growth, and motion. Such interactions are often complex and impossible to measure quantitatively. Thus, qualitative models, such as boolean networks and graph based methods, are useful tools when trying to make sense of very coarse measurements indicating a correlation or static relationship among different species. When dynamic data are available, it is possible to build qualitative ordinary differential equation models. Rather than choosing specific functional forms to model species interactions (such as Hill functions or polynomial terms), one can just make

general assumptions on the sign, trend and boundedness of said interactions. While such models are clearly not amenable to data fitting, they still allow us to reach useful analytical conclusions on the potential dynamic behaviors of a system.

The general class of qualitative biological models we consider are ordinary differential equations whose terms belong to four different categories:

$$\dot{x}_i(t) = \sum_{j \in \mathcal{A}_i} a_{ij}(x)x_j - \sum_{h \in \mathcal{B}_i} b_{ih}(x)x_h + \sum_{s \in \mathcal{C}_i} c_{is}(x) + \sum_{l \in \mathcal{D}_i} d_{il}(x). \quad (2.2)$$

Variables  $x_i$ ,  $i = 1, \dots, n$  are concentrations of species. The different terms in Eq. (2.2) are associated with a specific biological and physical meaning. Terms  $a_{ij}(x)x_j$  are associated with production rates of reagents; typically, these functions are assumed to be polynomial in their arguments; similarly, terms  $b_{ih}(x)x_h$  model degradation or conversion rates and are also likely to be polynomial in practical cases. Finally, terms  $c(\cdot)$  and  $d(\cdot)$  are associated with monotonic nonlinear terms, respectively non-decreasing and non-increasing; these terms are a qualitative representation of Michaelis-Menten or Hill functions [20].

Sets  $\mathcal{A}_i$ ,  $\mathcal{B}_i$ ,  $\mathcal{C}_i$ ,  $\mathcal{D}_i$  denote the subsets of variables affecting  $x_i$ . In general, more than one species can participate in the same term affecting a given variable. For instance one may have an interaction  $2 \rightarrow 1$  influenced also by species  $x_3$ :  $a_{12}(x_1, x_3)x_2$ . (The alternative notation choice,  $a_{13}(x_1, x_2)x_3$  would be possible.) To keep our notation simple, we do not denote external inputs with a different symbol. Inputs can be easily included as dynamic variables  $\dot{x}_u = w_u(x_u, t)$  which are not affected by other states and have the desired dynamics.

## 2.2.1 General Assumptions

We denote with  $\tilde{x}_i = [x_1 \ x_2 \ \dots \ x_{i-1} \ x_{i+1} \ \dots \ x_n]$  the vector of  $n - 1$  components complementary to  $x_i$  (e.g. in  $\mathbb{R}^4$   $\tilde{x}_2 = [x_1 \ x_3 \ x_4]$ ). Then  $f(x) = (\tilde{x}_j, x_j)$  for all  $j$ . In the remainder of this chapter, we assume that system (2.2) satisfies the following assumptions:

**A 1 (Smoothness)** Functions  $a_{ij}(\cdot)$ ,  $b_{ih}(\cdot)$ ,  $c_{is}(\cdot)$  and  $d_{il}(\cdot)$  are nonnegative, continuously differentiable functions.

**A 2** Terms  $b_{ij}(x)x_j = 0$ , for  $x_i = 0$ . This means that either  $i = j$  or  $b_{ij}(\tilde{x}_i, 0) = 0$ .

**A 3** Functions  $b_{ij}(x)x_j$  and  $a_{ih}(x)x_h$ , are strictly increasing in  $x_j$  and  $x_h$  respectively.

**A 4 (Saturation)** Functions  $c_{is}(\tilde{x}_s, x_s)$  are nonnegative and non-decreasing in  $x_s$ , while  $d_{il}(\tilde{x}_l, x_l)$  are nonnegative and, respectively, non-decreasing in  $x_l$ . Moreover  $c_{is}(\tilde{x}_s, \infty) > 0$  and  $d_{il}(\tilde{x}_l, 0) > 0$ . Moreover they are globally bounded.

In view of the nonnegativity assumptions and Assumption 2, our general model (2.2) is a nonlinear positive system and its investigation will be restricted to the positive orthant. We note that reducing dynamic interactions to a form  $b_{ij}(x)x_j$  and  $a_{ih}(x)x_h$  is always possible under mild assumptions: for instance, if species  $j$  affects species  $i$  with a monotonic functional term  $f_{ij}(\tilde{x}_j, x_j)$ , if such term has a locally bounded derivative, with  $f(\tilde{x}_i, 0) = 0$ , it can always be rewritten as:  $f_{ij}(x) = (f_{ij}(x)/x_j)x_j = a_{ij}(x)x_j$  (see [14], Sect. 2.1). Using the general class of models (2.2) and assumptions A1–A4 as a working template for analysis, we will focus on a series of paradigmatic dynamic behaviors which can be structurally identified or ruled out in example systems of interest.

### 2.2.2 Glossary of Properties

The structural analysis of system (2.2) can be greatly facilitated whenever it is legitimate to assume that functions  $a, b, c, d$  have certain properties such as positivity, monotonicity, boundedness and other functional characteristics that can be considered “qualitative and structural properties” [15]. Through such properties, we can draw conclusions on the dynamic behaviors of the considered systems without requiring specific knowledge of parameters and without numerical simulations. However, it is clear that our approach requires more information than other methods, such as boolean networks and other graph-based frameworks.

For the reader’s convenience, a list of possible properties and their definitions is given below, for functions of a scalar variable  $x$ .

**P 1**  $f(x) = \text{const} \geq 0$  is nonnegative-constant.

**P 2**  $f(x) = \text{const} > 0$  is positive-constant.

**P 3**  $f(x)$  is sigmoidal: it is non-decreasing,  $f(0) = f'(\infty) = 0$ , if  $0 < f(0) < \infty$  and its derivative has a unique maximum point,  $f'(x) \leq f'(\bar{x})$  for some  $\bar{x} > 0$ .

**P 4**  $f(x)$  is complementary sigmoidal: it is non-increasing,  $0 < f(0), f'(\infty) = 0$ ,  $f(\infty) = 0$  and its derivative has a unique minimum point. In simple words,  $f$  is a CSM function iff  $f(0) - f(x)$  is a sigmoidal function.

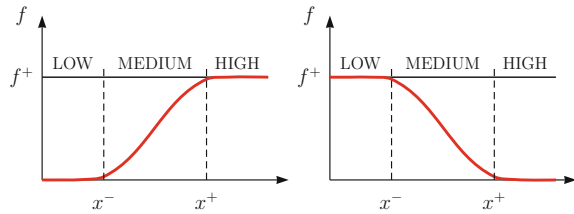
**P 5**  $f(x)$  is constant-sigmoidal, the sum of a sigmoid and a positive constant.

**P 6**  $f(x)$  is constant-complementary-sigmoidal, the sum of a complementary sigmoid and a constant.

**P 7**  $f(x)$  is increasing-asymptotically-constant:  $f'(x) > 0$ ,  $0 < f(\infty) < \infty$  and its derivative is decreasing.

**P 8**  $f(x)$  is decreasing-asymptotically-null:  $f'(x) < 0$ ,  $f(\infty) = 0$  and its derivative is increasing.

**Fig. 2.1** Cropped sigmoids and complementary sigmoids



**P 9**  $f(x)$  is decreasing-exactly-null:  $f'(x) < 0$ , for  $x < \bar{x}$  and  $f(x) = 0$  for  $x \geq \bar{x}$  for some  $\bar{x} > 0$ .

**P 10**  $f(x)$  is increasing-asymptotically-unbounded:  $f'(x) > 0$ ,  $f(\infty) = +\infty$ .

As an example, the terms  $d(\cdot)$  and  $c(\cdot)$  in general are associated with Hill functions, which are *sigmoidal* and *complementary sigmoidal* functions. In some cases it will be extremely convenient to introduce assumptions which are mild in a biological context but assure a strong simplification of the mathematics. One possible assumption is that a sigmoid or a complementary sigmoid is cropped (Fig. 2.1). A cropped sigmoid is exactly constant above a certain threshold  $x^-$  and exactly null below another threshold  $x^+$ . A cropped complementary sigmoid is exactly null above  $x^-$  and exactly constant below  $x^+$ .

These assumptions extend obviously to multivariable functions just by considering one variable at the time. For instance  $f(x_1, x_2)$  can be a sigmoid in  $x_1$  and decreasing in  $x_2$ .

### 2.2.3 Network Graphs

Building a dynamical model for a biological system is often a long and challenging process. For instance, to reveal dynamic interactions among a pool of genes of interest, biologists may need to selectively knockout genes, set up micro RNA assays, or integrate fluorescent reporters in the genome. The data derived from such experiments are often noisy and uncertain, which implies that also the estimated model parameters will be uncertain. However, *qualitative trends* can be reliably assessed in the dynamic or steady state correlation of biological quantities. Graphical representations of such qualitative trends are often used by biologists, to provide intuition regarding the network main features.

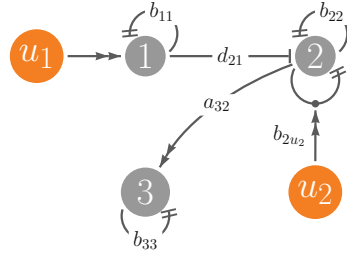
Building on the general model (2.2), we can associate species to nodes of a graph, and different qualitative relationships between species with different types of arcs: terms  $a$ ,  $b$ ,  $c$  and  $d$  can be represented as arcs having different end-arrows, as shown in Fig. 2.2.

These graphs can be immediately constructed, by knowing the correlation trends among the species of the network, and serve as a support for the construction and

**Fig. 2.2** Arcs associated to the different terms of our general model (2.2), and example graph



**Fig. 2.3** Graph corresponding to the transcriptional repression example in Sect. 2.1.1



analysis of a dynamical model. For simple networks, these graphs may facilitate structural robustness analysis.

Our main objective is to show that, at least for reasonably simple networks, structural robust properties can be investigated with simple analytical methods, without the need for extensive numerical analysis. We suggest a two stage approach:

- Preliminary screening: establish essential information on the network structure, recognizing which properties (such as P1–P10) pertain to each link.
- Analytical investigation: infer robustness properties based on dynamical systems tools such as Lyapunov theory, set invariance and linearization.

## 2.2.4 Example, Continued: Transcriptional Repression

The model for the transcriptional repression system in Eq. (2.1) [19] can be recast in the general class of models (2.2), and we can immediately draw the corresponding graph (Fig. 2.3).

$$\begin{aligned} \dot{x}_1 &= u_1 - b_{11}x_1, \\ \dot{x}_2 &= d_{21}(x_1) - b_{22}x_2 - b_{2u_2}x_2 u_2, \\ \dot{x}_3 &= a_{32}x_2 - b_{33}x_3. \end{aligned} \quad (2.3)$$

Terms  $a_{ij}$  capture first order production rates;  $b_{ih}$  capture first order degradation rates. Term  $d_{21}(x_1)$  is our general substitute for the Hill function [19, 20]; we assume it is a decreasing function with null derivative at the origin, whose second derivative has a single zero (flexus), and it is negative on the left of the zero and positive on the right (such as  $1/(1+x_1^n)$ ,  $n > 1$ ).



## 2.3 Robustness and Structural Properties

We now clarify the concepts of robustness and structural properties and their relations.

**Definition 1** Let  $\mathcal{C}$  be a class of systems and  $\mathcal{P}$  be a property pertaining such a class. Given a family  $\mathcal{F} \subset \mathcal{C}$  we say that  $\mathcal{P}$  is robustly verified by  $\mathcal{F}$ , in short robust, if it is satisfied by each element of  $\mathcal{F}$ .

Countless examples can be brought about families  $\mathcal{F}$  and candidate properties. Stability of equilibria, for instance, is one of the most investigated structural properties [2, 13, 21].

When we say *structural property* we refer to the properties of a family  $\mathcal{F}$  whose “structure” has been specified. In our case, the structure of a system is the fact that it belongs to the general class (2.2), thus it satisfies assumptions 1–3, and it enjoys properties in the set P1–P8.

A *realization* is any system with assumed structure and properties achieved by specific functions which satisfy these assumptions. The set of all realization is a *class*. For instance, going back to the transcriptional repression example, the dynamical system:

$$\begin{aligned}\dot{x}_1 &= u_1 - 2x_1, \\ \dot{x}_2 &= \frac{1}{1 + x_1^n} - x_2 - 2x_2u_2, \\ \dot{x}_3 &= 2x_2 - 2x_3,\end{aligned}$$

is a realization of the class represented by system (2.3).

**Definition 2** A property  $\mathcal{P}$  is structural for a class  $\mathcal{C}$ , if any realization satisfies  $\mathcal{P}$ .

Note that demonstrating a structural property for a system is harder than proving that it does not hold (the latter typically only requires to show the existence of a system which exhibits the considered structure but does not satisfy the property). For example, consider matrices:

$$A_1 = \begin{bmatrix} -a & b \\ -c & -d \end{bmatrix} \quad A_2 = \begin{bmatrix} -a & b \\ c & -d \end{bmatrix}$$

with  $a, b, c$  and  $d$  positive real parameters. To show that  $A_1$  is structurally stable one has to show that its eigenvalues have negative real part, (in this case, a simple proof). Conversely to show that  $A_2$  is not structurally stable, it is sufficient to find a realization which is not stable, such as  $a = 1$   $b = 1$   $c = 2$  and  $d = 1$ .

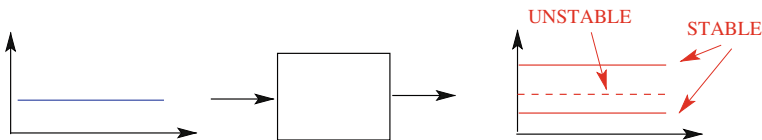


Fig. 2.4 Sketch of a bistable system

## 2.4 Paradigmatic Structural Properties

We introduce an overview of properties particularly relevant in systems and synthetic biology. Through simple examples, we highlight how our general approach can be used to determine analytically the structural nature of such properties.

### 2.4.1 Multistationarity

A multistationary system is characterized by the presence of several possible equilibria. Of particular interest are those systems in which there are three equilibria, of which two are stable and one unstable, *i.e.*, the system is bistable.

We consider a simple example of a multistationary system (Fig. 2.4):

$$\begin{aligned}\dot{x}_1 &= x_0 + c_{12}(x_2) - b_{11}x_1 \\ \dot{x}_2 &= a_{21}x_1 - b_{22}x_2\end{aligned}\tag{2.4}$$

with  $b_{11}$ ,  $b_{22}$  and  $a_{21}$ , positive constants, and with  $c_{12}(x_2)$  a (non-decreasing) sigmoidal function. We assume  $x_0 \geq 0$ . The following proposition holds:

**Proposition 1** *For  $x_0$  small enough and for  $b_{11}b_{22}/a_{21}$  small enough, system (2.4) has three equilibria, two stable and one unstable. Conversely, for  $x_0$  large or  $b_{11}b_{22}/a_{21}$  large the system admits a unique, stable equilibrium.*

**Explanation.** Setting  $\dot{x}_1 = 0$  and  $\dot{x}_2 = 0$  we find the equilibria as the roots of the following equation:

$$c_{12}(x_2) + x_0 = \frac{b_{11}b_{22}}{a_{21}}x_2$$

From Fig. 2.5, it is apparent that if  $x_0$  is small and the slope of the line  $\frac{b_{11}b_{22}}{a_{21}}x_2$  is small, there must be three intersections. Conversely, there is a single intersection for either  $x_0$  or  $\frac{b_{11}b_{22}}{a_{21}}$  large.  $\square$

If three intersections (points A, B, C in Fig. 2.5) are present, there are two stable points A and B and one unstable. This can be seen by inspecting the Jacobian:

$$J = \begin{bmatrix} -b_{11} & c'_{12}(\bar{x}_2) \\ a_{21} & -b_{22} \end{bmatrix},$$

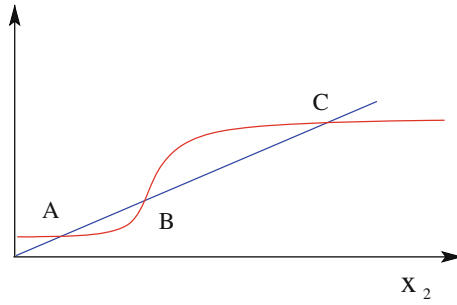


Fig. 2.5 Sketch of the nullclines for system (2.4)

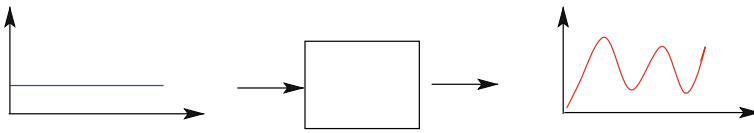


Fig. 2.6 Schematic representation of oscillatory behavior

whose characteristic polynomial is:

$$p(s) = s^2 + (b_{11} + b_{22})s + b_{11}b_{22} - a_{21}c'_{12}(\bar{x}_2).$$

This second order polynomial is stable if  $b_{11}b_{22} - a_{21}c'_{12}(\bar{x}_2) > 0$  or

$$c'_{12}(x_2) < \frac{b_{11}b_{22}}{a_{21}},$$

namely the slope of the sigmoidal function must be smaller than the slope of the line  $b_{11}b_{22}/a_{21}$ . This is the case of points A and C, while the condition is violated at point B.

### 2.4.2 Oscillations

Oscillations in molecular and chemical networks are a well-studied phenomenon (see, for instance [22]). Periodicity in molecular concentrations underlies cell division, development, and circadian rhythms. One of the first examples considered in the literature is the well known Lotka Volterra predator-prey system, whose biochemical implementation has been studied and attempted in the past [23, 24]. In our general setup, the Lotka Volterra model is (Fig. 2.6):

$$\begin{aligned}\dot{x}_1 &= a_{11}x_1 - b_{12}(x_2)x_1 \\ \dot{x}_2 &= a_{21}(x_1)x_2 - b_{22}x_2,\end{aligned}$$

where all functions are strictly increasing and asymptotically unbounded in all arguments. The system admits a single non-trivial equilibrium, the solution of equations:

$$\begin{aligned}0 &= a_{11} - b_{12}(x_2) \\ 0 &= a_{21}(x_1) - b_{22}.\end{aligned}$$

The Jacobian of this system at the unique equilibrium is:

$$J = \begin{bmatrix} 0 & -b'_{12}(x_2)x_1 \\ a'_{21}(x_1)x_2 & 0 \end{bmatrix}.$$

This matrix clearly admits pure imaginary eigenvalues for any realization of the functional terms. Thus, oscillations are a structural property.

In second order systems, sustained oscillations require the presence of a positive self loop (autocatalytic reactions) represented in this case by the  $a_{11}$  term.

To achieve oscillations without a positive loop reaction, the system must be of at least third order. For instance the following model

$$\begin{aligned}\dot{x}_1 &= x_{10}d_{13}(x_3) - b_{11}x_1 \\ \dot{x}_2 &= a_{21}x_1 - b_{22}x_2, \\ \dot{x}_3 &= a_{32}x_2 - b_{33}x_3,\end{aligned}\tag{2.5}$$

where  $d_{13}(x_3)$  is a complementary sigmoid and the constant are positive, is a candidate oscillator. Term  $x_{10}$  is an external input which catalyzes the production  $d_{13}(x_3)$ .

**Proposition 2** *System (2.5) admits a unique equilibrium. If the minimum value of the slope  $d'_{13}(x_3)$  is sufficiently large, there exists an interval (possibly unbounded from above) of input values  $x_{10}$  inducing an oscillatory transition to instability.*

**Explanation** The unique equilibrium point can be derived by the conditions  $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$ :

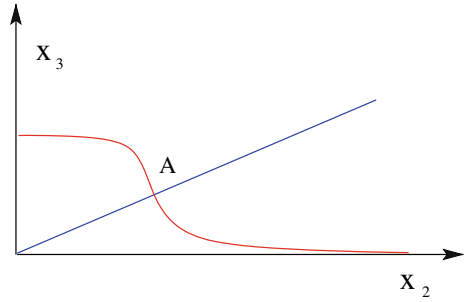
$$x_{10}d_{13}(x_3) = \frac{b_{11}b_{22}b_{33}}{a_{21}a_{32}}x_3.$$

Figure 2.7 shows the qualitative trend of the nullclines above, and clearly highlights that they admit a single intersection.

Assume that the slope in the intersection point  $A$  is large. The Jacobian of the system at this equilibrium point is

$$J = \begin{bmatrix} -b_{11} & 0 & -\mu \\ a_{21} & -b_{22} & 0 \\ 0 & b_{32} & -b_{33} \end{bmatrix}, \quad \mu = -x_0d'_{13}(\bar{x}_3) > 0.$$

**Fig. 2.7** Qualitative trend of the nullclines for system (2.5).



The corresponding characteristic polynomial is

$$p(s) = (s + b_{11})(s + b_{22})(s + b_{33}) + a_{21}a_{32}\mu = s^3 + p_2s^2 + p_1s + p_0 + a_{21}a_{32}\mu.$$

This polynomial has a pair of complex conjugate roots with positive real part, as it can be inferred from the Ruth–Hurwitz table:

+	1	$p_1$
+	$p_2$	$p_0 + a_{21}a_{32}\mu$
?	$(p_1p_2 - a_{21}a_{32}\mu)/p_2$	
+	$a_{21}a_{32}\mu$	

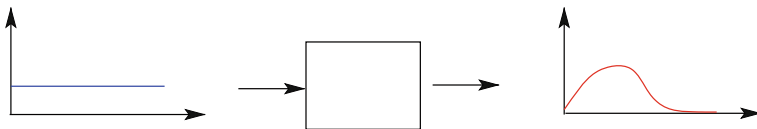
for large  $\mu$  there are two sign in the first column of the table, which means that there are two unstable roots. These roots cannot be real because the polynomial coefficients are all positive, so unstable roots must be complex conjugate.

In general, we can say there is an “interval” in parameter space in which oscillations are admissible: for  $x_0$  small, the intersection occurs in a region where the slope of  $\mu = -x_0d'_{13}(\bar{x}_3)$  is small, thus there are no changes in the Routh-Hurwitz table and the system is stable.  $\square$

Note that it is not necessarily true that for large  $x_0$  the system is unstable; in addition, the instability interval of  $x_0$  may be bounded. In fact, the equilibrium  $\bar{x}_3$  increases for large  $x_0$ , but it may transition to a region where  $d'_{13}$  is very small, compensating for the increase of  $x_0$ .

### 2.4.3 Adaptation

A system is adaptive if, when perturbed by a persistent input signal, its output always reverts to a neighborhood of its value prior to the perturbation, in general after a transient [1, 25, 26]. A sketch of this behavior is in Fig. 2.8. Adaptation is said to be perfect if the system’s output reverts to its exact value prior to the perturbation.



**Fig. 2.8** System capable of adaptation

For small perturbations, linearization analysis suggests that adaptation requires the presence of a zero in the system's transfer function. If the system includes a feedback loop, then the presence of a pole at the origin (integrator) is required [25, 26]. Establishing criteria to detect a system's capability for adaptation is thus simple. Consider the system:

$$\dot{x}_1 = -b_{21}(x_1)x_2 + x_0, \quad (2.6)$$

$$\dot{x}_2 = a_{12}x_1 - b_{22}x_2 + u. \quad (2.7)$$

We assume all the constants are positive, and that function  $b_{21}(x_1)$  is a cropped sigmoid, namely it is strictly increasing and exactly positive constant above a certain threshold. Term  $x_0$  is a constant, and  $u \geq 0$  is a perturbing input.

**Proposition 3** *If  $x_0$  is sufficiently large and  $u = 0$ , then system (2.6) has a stable equilibrium point. Taking  $y = x_2$  as the system's output, perfect adaptation is achieved with respect to constant perturbations on  $u > 0$ .*

**Explanation.** For  $u = 0$  the equilibrium conditions are  $b_{21}(x_1)x_2 = x_0$  and  $a_{12}x_1 - b_{22}x_2 = 0$ . Therefore the equilibrium  $\bar{x}_1$  can be expressed as the solution of:

$$b_{21}(x_1) \frac{a_{12}}{b_{22}} x_1 = x_0. \quad (2.8)$$

For  $x_0$  suitably large,  $\bar{x}_1$  increases until it falls in the range where  $b_{21}$  (a cropped sigmoid) is constant, thus  $b_{21}(x_1) = b_{21}(\infty)$ , and  $b'_{21}(x_1) = 0$ .

In this range, the linearized system is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} \begin{bmatrix} 0 & -b_{21}(x_1) \\ a_{21} & -b_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \quad y = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

with output  $y(t) = x_2(t)$ . The state matrix is a stable matrix, with characteristic polynomial  $p(s) = s^2 + b_{22}s + b_{21}(x_1)a_{21}$ . The transfer function is  $w(s) = s/p(s)$ , has a zero at the origin and thus the system locally exhibits perfect adaptation.

If  $u > 0$  increases as a step input, after a transient the output  $x_2$  returns to its original value  $\bar{x}_2$  prior to the perturbation. However, the equilibrium of  $x_1$  increases to a new value such that  $\bar{a}_{12}\bar{x}_1 = b_{22}\bar{x}_2 + u$ .  $\square$

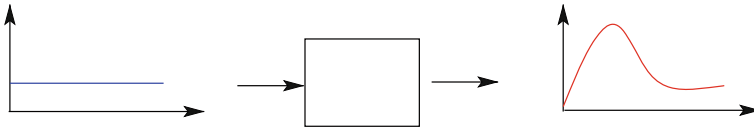


Fig. 2.9 System presenting a spiking behavior

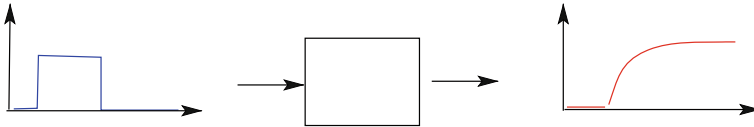


Fig. 2.10 System presenting a persistent response

### 2.4.4 Spiking and Persistency: The MAPK Network as a Case Study

Spiking is a phenomenon observed in several molecular networks, in which a system subject to a step input grows rapidly and subsequently undergoes a relaxation, as sketched in Fig. 2.9. The relaxation brings the system to a new equilibrium, distinct from the equilibrium prior to the input stimulation.

Persistency is closely related to bistability: it occurs when a transient input variation causes the system to switch its output to a new value, which persists upon removal of the input, as shown in Fig. 2.10.

#### 2.4.4.1 A Qualitative Model of the MAPK Pathway

Experiments show that the mitogen-activated protein kinase (MAPK) pathway in PC12 rat neural cells exhibits dynamic behaviors that depend on the growth factor they are exposed to as an input. The response to Epidermal Growth Factor (EGF) is a spike followed by a relaxation, while the response to Nerve Growth Factor (NGF) is persistent. In the latter case, the system can be driven to a new state, which persists after the stimulus has vanished. Ultimately, these dynamic behaviors correspond to different cell fates: EGF stimulation induces proliferation, while NGF stimulation induces differentiation. The biochemical mechanisms responsible for the different input-dependent dynamic response are still unclear. One hypothesis is that each input generates a specific interaction topology among the kinases. Starting from experimental results that support this hypothesis [27], in our previous work we considered the two network topologies, and we derived and analyzed qualitative models which exhibit structural properties [28]. Here we use a simplified, third order model for the pathway. We refer the reader to [28] for a more detailed model and its derivation. In our reduced order model, we neglect double-phosphorylation dynamics, and model the active concentration of each MAPK protein with a single state variable. We also neglect mass conservation assumptions regarding the total

amount of MAPK protein [13, 16].

$$\text{MAP3K: } \dot{x}_1 = u(x_3, x_0) - b_{11}x_1 \quad (2.9)$$

$$\text{MAP2K: } \dot{x}_2 = c_{21}(x_1) - b_{22}x_2 \quad (2.10)$$

$$\text{MAP1K: } \dot{x}_3 = c_{32}(x_2) - b_{33}x_3 \quad (2.11)$$

$$\text{Output: } y = x_3 \quad (2.12)$$

We assume:  $c_{21}$  and  $c_{32}$  are strictly increasing asymptotically constant, *i.e.*  $c_{21}(\infty) = \hat{c}_{21} < \infty$   $c_{32}(\infty) = \hat{c}_{32} < \infty$ , and null at the origin  $c_{21}(0) = c_{32}(0) = 0$ . Terms  $b_{ii}$  are positive constants. In essence, this model captures the fact that each protein in the cascade is activated by its predecessor in the chain; in the absence of term  $u(x_3, x_0)$ , the system would be an open loop, monotonic cascade [12]. Term  $u(x_3, x_0)$  is a feedback term modulated by an external input  $x_0$ , and we consider two cases:

EGF  $u = a_{10}(x_3)x_0$ , where  $a_{10}(x_3)$  is a complementary sigmoid, exactly constant below a threshold  $\eta$  and exactly null over a threshold  $\xi$ . This configuration is characterized by the presence of a negative feedback loop.

NGF  $u = a_{10}(x_3) + x_0$ , where  $a_{10}(x_3)$  is a sigmoid, exactly null below a threshold  $\eta$  and exactly constant over a threshold  $\xi$ . This configuration is characterized by the presence of a positive feedback loop.

Under these assumptions, we show that in the EFG configuration the output exhibits a spike, while in the NGF configuration the output is persistent.

### 2.4.5 The EGF-Induced Pathway and Its Spiking Behavior

The system in this configuration admits a single equilibrium; this can be shown as for the third order oscillator model (2.5).

Consider  $c_{21}(\infty) = \hat{c}_{21}$ ,  $c_{32}(\infty) = \hat{c}_{32}$ , the saturation value. Let  $\hat{x}_2 = \hat{c}_{21}/b_{22}$  be the corresponding ‘‘saturation’’, limit value of  $x_2$ . Let, in turn,

$$\hat{x}_3 = c_{32}(\hat{x}_2)/b_{33}$$

be the limit value of  $x_3$ . For large, increasing values of the input  $x_0$ , the variable  $\hat{x}_1$  increases and the equilibrium values of  $x_2$  and  $x_3$  approach  $\hat{x}_2$  and  $\hat{x}_3$ . The following proposition holds:

**Proposition 4** *Assume that the limit value for  $x_3$  is  $\hat{x}_3 > \xi$ . Then, for  $x_0$  constant sufficiently large, and for  $x_i(0) = 0$ , we have: (a) First,  $x_3$  grows arbitrarily close to  $\hat{x}_3$ . (b) Subsequently,  $x_3$  relaxes below  $\xi$ .*

*Proof* Since  $a_{10}(x_3)$  is constant for a small values of  $x_3$ , if  $x_0$  is large then by continuity  $x_1$  can grow arbitrarily large in an arbitrarily small amount of time  $\tau > 0$ .



Then, considering the time interval  $[\tau, T]$  where  $T$  is arbitrarily large, and given an arbitrary  $\mu > 0$ , by picking  $x_1(\tau)$  sufficiently large we can guarantee:

$$x_1(t) \geq \mu \quad \text{for } t \in [\tau, T]. \quad (2.13)$$

In fact, we have  $\dot{x}_1 \geq -b_{11}x_1$ , thus  $x_1(t) \geq x_1(\tau)e^{-b_{11}t}$  on  $[\tau, T]$ ; therefore, picking a large initial value  $x_1(\tau)$ , equation (2.13) is verified. Thus, we can guarantee that variables  $x_2$  and  $x_3$  have values arbitrarily close to the upper limit  $\hat{x}_2$  and  $\hat{x}_3$ , being  $\mu$  and  $T$  arbitrarily large.

If  $x_3$  increases, at some point in time the condition  $a_{10}(x_3) = 0$  is met. This “switches off” the first variable, whose dynamics become:  $\dot{x}_1 = -b_{11}x_1$ , thus  $x_1$  starts decreasing; variables  $x_2$  and  $x_3$  follow the same pattern. These concentrations decrease until  $x_3 \leq \xi$ .  $\square$

### 2.4.6 The NGF-Induced Pathway Is an Example of Persistent Network

Let us now define  $a_{10}(\infty) = \bar{a}_{10}$  as a saturation value. If  $\bar{x}_3$  is greater than the threshold  $\xi$ , then  $a_{10}(\bar{x}_3) = \bar{a}_{10}$ ; then, for  $x_0 = 0$  we can find the equilibria from the following conditions:

$$0 = a_{10} - b_{11}\bar{x}_1, \quad (2.14)$$

$$0 = c_{21}(\bar{x}_1) - b_{22}\bar{x}_2, \quad (2.15)$$

$$0 = c_{32}(\bar{x}_2) - b_{33}\bar{x}_3, \quad (2.16)$$

which yield  $\bar{x}_1 = \bar{a}_{10}/b_{11}$ ;  $\bar{x}_2 = c_{21}(\bar{x}_1)/b_{22}$ ;  $\bar{x}_3 = c_{32}(\bar{x}_2)/b_{33}$ . The assumption  $\bar{x}_3 > \xi$  means that the positive feedback given by the term  $\bar{a}_{10}$  is able to sustain this positive equilibrium.

Now consider the case where the input  $x_0$  becomes arbitrarily large. Thus,  $\bar{x}_1$  becomes arbitrarily large. Defining  $\hat{c}_{21} = c_{21}(\infty)$ , we find the corresponding limit values for the steady states:  $\hat{x}_2 = \hat{c}_{21}/b_{22}$  and  $\hat{x}_3 = c_{32}(\hat{x}_2)/b_{33}$ . It is immediate that  $\hat{x}_1 \geq \bar{x}_1$ ,  $\hat{x}_2 \geq \bar{x}_2$ ,  $\hat{x}_3 \geq \bar{x}_3$ , because the “hat” equilibrium values are achieved by means of an arbitrarily large input  $x_0$ , while the “bar” values are achieved by the bounded input  $\bar{a}_{10}$ .

**Proposition 5** *Assume that  $\bar{x}_3 > \xi$  and that the previous inequalities are strict:  $\hat{x}_1 > \bar{x}_1$ ,  $\hat{x}_2 > \bar{x}_2$ ,  $\hat{x}_3 > \bar{x}_3$ . Then, for  $x_i(0) = 0$  the following happens:*

- (a) *If  $x_0$  is constant and sufficiently large, and it is applied for a sufficiently long time interval  $[0, T]$ , then  $x_3$  grows arbitrary close to  $\hat{x}_3$ .*
- (b) *If, after time  $T$ , the input signal  $x_0$  is eliminated ( $x_0 = 0$ ), then  $x_3$  remains above  $\xi$ .*
- (c) *Finally,  $x_3$  converges to  $\bar{x}_3$  from above.*

*Proof* We have seen that when  $x_0 = 0$ ,  $\bar{x}_1, \bar{x}_2, \bar{x}_3$  are admissible equilibria of the system. Exactly as done in the EGF-driven network example, we can show that for a sufficiently large input  $x_0$ , variables  $x_1$ ,  $x_2$  and  $x_3$  can grow arbitrarily close to  $\hat{x}_1$ ,  $\hat{x}_2$  and  $\hat{x}_3$ , above  $\bar{x}_1$ ,  $\bar{x}_2$  and  $\bar{x}_3$ .

We only need to show that if all  $x_i(t)$  grow above the corresponding  $\bar{x}_i$ , then they will not reach values below  $\bar{x}_i$  after  $x_0$  is removed.

We begin by defining the new variables  $z_i = x_i - \bar{x}_i$ ; then,  $\dot{z}_i = \dot{x}_i$  given by equations (2.9)–(2.11). After  $x_0$  is removed, the input is  $a_{10}(x_3)$ ; in addition, since we assume  $x_3 \geq \bar{x}_3 \geq \xi$  (so  $z_3 \geq 0$ ), we have  $a_{10}(x_3) = \bar{a}_{10}$ . If we consider also the steady state equations (2.14)–(2.16), we get

$$\dot{z}_1 = -b_{11}z_1 \quad (2.17)$$

$$\dot{z}_2 = c_{21}(z_1 + \bar{x}_1) - c_{21}(\bar{x}_1) - b_{22}z_2 \quad (2.18)$$

$$\dot{z}_3 = c_{32}(z_2 + \bar{x}_2) - c_{32}(\bar{x}_2) - b_{33}z_3 \quad (2.19)$$

This is a positive system in the  $z$  variables. Because we assumed that at some point  $z_i(\tau) > 0$  (prior to the removal of  $x_0$ ), we can immediately see that this situation is permanent.

To prove convergence, note that  $z_1$  goes to zero in view of Eq. (2.17). Then  $c_{21}(z_1 + \bar{x}_1) - c_{21}(\bar{x}_1)$  goes to 0, so  $z_2$  converges to 0. For the same reason,  $z_3$  converges to 0.  $\square$

## 2.5 Structural Boundedness and Stability

Our qualitative modeling framework is generally described by Eq. (2.2):

$$\dot{x}_i(t) = \sum_{j \in \mathcal{A}_i} a_{ij}(x)x_j - \sum_{h \in \mathcal{B}_i} b_{ih}(x)x_h + \sum_{s \in \mathcal{C}_i} c_{is}(x) + \sum_{l \in \mathcal{D}_i} d_{il}(x).$$

The general assumptions we made on functions  $a$ ,  $b$ ,  $c$ , and  $d$  guarantee non-negativity of the states, which is a required feature to meaningfully model concentrations of molecules. Another important feature of most biochemical system models is boundedness of their states (possibly with the exception of pathological cases). In the following, we outline additional assumptions and consequent results regarding structural boundedness of the solutions to our general model (2.2).

### 2.5.1 Structural Boundedness

Consider the case in which states in model (2.2) are dissipative, i.e. the dynamics of each variable include a degradation term  $-b_{ii}(x)x_i$ . We also assume that

$$b_{ii}(x) > \beta_i > 0.$$

Obviously, this property alone does not assure the global boundedness of the solution. However, if no unbounded  $a$ -terms were present, it would be simple to show that the solutions are globally bounded.

Let us assume that each  $a_{ij}(x)$  term is bounded by a positive constant  $0 \leq a_{ij}(x) < \bar{a}_{ij}$ . Then, we ask under what conditions we can assure structural boundedness of the solutions. We build a graph  $G(\mathcal{A})$  associated with the  $a_{ij}$  terms, where there is a directed arc from node  $j$  to node  $i$  for every term  $a_{ij}$ . Then, the following theorem holds.

**Theorem 1** *The system solution is structurally globally bounded for any initial condition  $x(0) \geq 0$  if and only if  $G(\mathcal{A})$  has no cycles (including self-cycles) including  $a_{ii}$  terms.*

In other words, structural boundedness is guaranteed if and only if there is no auto-catalysis in the system.

*Proof* We first show that the condition is structurally necessary. Assume, *ab absurdo*, that there is a cycle which includes a term  $a_{ij}$ . Without restriction assume that the cycle is formed by the first  $r$  nodes  $1, 2, \dots, r$ , forming a sequence  $a_{12}, a_{23}, \dots, a_{r1}$ ; also, assume that each term  $a_{ij}$  is lower bounded by a constant  $\kappa$ . We finally assume that the sum of all  $b_{ik}$  terms appearing in the first  $r$  equations is upper bounded by  $\eta$ :

$$\sum_{i=1}^r \sum_{k \in \mathcal{B}_i} b_{ik} \leq \eta.$$

Consider the Lyapunov-like function:

$$V(x_1, x_2, \dots, x_r) = x_1 + x_2 + \dots + x_r,$$

and its derivative

$$\begin{aligned} \dot{V} &= \sum_{i=1}^r \dot{x}_i \geq \sum_{i=1}^r \left[ a_{i,i+1} x_{i+1} - \sum_{k \in \mathcal{B}_i} b_{ik} x_k \right] \geq \sum_{i=1}^r a_{i,i+1} x_{i+1} - \eta \sum_{i=1}^r x_i \\ &\geq (r\kappa - \eta) \sum_{i=1}^r x_i = (r\kappa - \eta)V. \end{aligned}$$

Then, if  $\eta < r\kappa$ ,  $V$  increases and the equilibrium is not stable. Thus, structural boundedness cannot hold.

Let us now consider the sufficiency part. If there are no cycles in  $G(\mathcal{A})$ , then there exists necessarily a node which is a root, i.e. its dynamics do not include  $a_{ij}$  terms.

Let us assume, without loss of generality, that node  $x_1$  does not have any  $a_{1j}$  term. Then:

$$\begin{aligned}\dot{x}_1 &= - \sum_{h \in \mathcal{B}_1} b_{ih}(x)x_h + \sum_{s \in \mathcal{C}_1} c_{1s}(x) + \sum_{l \in \mathcal{D}_1} d_{1l}(x) \\ &\leq -\beta_1 x_1 + \sum_{s \in \mathcal{C}_1} c_{1s}(x) + \sum_{l \in \mathcal{D}_1} d_{1l}(x)\end{aligned}$$

Since the  $c$  and  $d$  terms are bounded, then the solution  $x_1$  is bounded; without loss of generality, assume  $x_1 \leq \xi_1$ ,  $\xi_1 > 0$ .

If  $x_1$  is bounded, then all terms (if any) of type  $a_{k1}(x)x_1$  in other equations remain bounded:  $a_{k1}(x)x_1 \leq \bar{a}_{j1}\xi_1$ .

Let us consider the other nodes  $x_2, x_3, \dots, x_n$ . Since there are no cycles including  $a_{ij}$  terms, there is at least one variable whose equation has either no  $a$  terms, or has only  $a_{k1}(x)x_1$  terms from  $x_1$ , which are bounded. Let us assume node  $x_2$  fulfills this statement. Then:

$$\begin{aligned}\dot{x}_2 &= a_{i1}(x)x_1 - \sum_{h \in \mathcal{B}_2} b_{ih}(x)x_h - \sum_{h \in \mathcal{B}_2} b_{ih}(x)x_h + \sum_{s \in \mathcal{C}_2} c_{2s}(x) + \sum_{l \in \mathcal{D}_2} d_{2l}(x) \\ &\leq -\beta_2 x_2 + \bar{a}_{j1}\xi_1 + \sum_{s \in \mathcal{C}_2} c_{2s}(x) + \sum_{l \in \mathcal{D}_2} d_{2l}(x).\end{aligned}$$

The above inequality implies boundedness of the solution  $x_2$ .

The proof can be concluded recursively, by noticing that there must exist a new variable, say  $x_3$  whose equation includes either no  $a_{ij}$  terms or only bounded  $a_{3j}$  terms coming from  $x_1$  and  $x_2$ , and so on.  $\square$

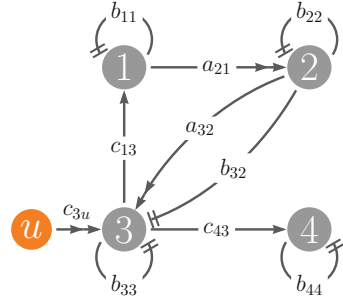
The following corollary holds.

**Corollary 1** *The solution to the general model (2.2) is bounded if and only there are no  $a_{ij}$  terms and all  $b_{ii}$  terms are lower bounded by a positive constant,  $b_{ii} > \beta_i$ .*

This corollary highlights that boundedness is structurally assured in systems where each species is degraded by terms of at least first order, and all the interaction terms are bounded.

*Example 1* As an example we consider the well known *lac* Operon genetic network. We will propose and analyze a qualitative model or class: the classical model proposed in [29] is a realization within this class. The state variables of our model are: the concentration of nonfunctional permease protein  $x_1$ ; the concentration of functional permease protein  $x_2$ ; the concentration of inducer (allolactose) inside the cell  $x_3$ , and the concentration of  $\beta$ -galactosidase  $x_4$ , a quantity that can be experimentally measured. The concentration of inducer external to the cell is here denoted as an input function  $u$ . A model for this system can be written in the following form (see [16] for details).

**Fig. 2.11** Graph of the lac operon network



$$\begin{aligned}
 \dot{x}_1 &= c_{13}(x_3) - b_{11}x_1, \\
 \dot{x}_2 &= a_{21}x_1 - b_{22}x_2, \\
 \dot{x}_3 &= a_{32}(u)x_2 - b_{32}(x_3)x_2 + c_{3u}u - b_{33}x_3, \\
 \dot{x}_4 &= c_{43}(x_3) - b_{44}x_4,
 \end{aligned} \tag{2.20}$$

where  $c_{13}(x_3) = f_1(x_3)$ ,  $b_{11} = \delta_1$ ,  $a_{21} = \beta_1$ ,  $b_{22} = \delta_2$ ,  $a_{32}(u) = f_2(u) =$ ,  $b_{32}(x_3) = f_3(x_3)$ ,  $c_{3u} = \beta_2$ ,  $b_{33} = \delta_3$ ,  $c_{43}(x_3) = \gamma f_1(x_3)$  and  $b_{44} = \delta_4$ . This corresponds to the network in Fig. 2.11.

We assume that  $c_{13}$  is *constant-sigmoidal*,  $a_{32}(u)$  and  $b_{32}(x_3)$  are *increasing-asymptotically-constant*, and the remaining functions  $a_{21}$ ,  $b_{11}$ ,  $b_{22}$  and  $b_{33}$  are *positive-constant*.

The arcs associated to  $a_{ij}$  terms in Fig. 2.11 do not form any cycles. Each node is dissipative, therefore the solution is structurally bounded.

The requirement of having no  $a_{ij}$  cycles can be strong, especially in chemical reaction networks [11]. However, the conditions in Theorem 1 are necessary and sufficient; we believe it is unlikely that stronger results can be found without assuming bounds on the dynamic terms.

Note that Theorem 1 only requires that bounds on the functional terms exist, while their specific values need not be known. If such bounds are known, we obtain less restrictive conditions. Note that model (2.2) can be written compactly as:

$$\dot{x}(t) = A(x(t))x(t) - B(x(t))x(t) + C(x(t)) + D(x(t)), \tag{2.21}$$

or as:

$$\dot{x}(t) = M(x(t))x(t) + C(x(t)) + D(x(t)), \tag{2.22}$$

where  $M(x(t)) = A(x(t)) - B(x(t))$ . If the elements of matrix  $M(x(t))$  are constrained in a closed (even better if compact) set,  $M(\cdot) \in \mathcal{M}$ , and if and if we can demonstrate exponential stability of the associated differential inclusion [30]

$$\dot{x} \in \mathcal{M}x,$$

then we can show the overall boundedness of the systems' solution. To prove boundedness it is convenient to exclude a neighborhood of the origin:  $\mathcal{N}_\nu = \{x : x_i \geq \nu\}$ .

**Theorem 2** *Assume that  $M(x) \in \mathcal{M}$  for  $x \in \mathcal{N}$  and assume that the differential inclusion is bounded and admits a positively homogeneous function  $V(x)$  as Lyapunov function*

$$\dot{V}(x) = \nabla V(x)Mx \leq -\gamma V(x)$$

for all  $M \in \mathcal{M}$ . Then the system solution is bounded.

*Proof* The proof is an immediate consequence of the fact that the trajectories of the original linear systems are a subset of the possible trajectories of the linear differential inclusions.

An exponentially stable differential inclusion has bounded solutions if perturbed by bounded terms

$$\dot{x} \in \mathcal{M}x + C + D$$

as in our case. □

*Example 2* Consider a biological network composed by two proteins  $x_1$  and  $x_2$ :

$$\begin{aligned}\dot{x}_1 &= +c_{10} + a_{12}(x_1)x_2 - b_{11}x_1, \\ \dot{x}_2 &= +c_{20} - b_{21}(x_2)x_1 - b_{22}x_2.\end{aligned}$$

In this model, we suppose that both  $x_1$  and  $x_2$  are produced in active form at some constant rates (terms  $c_{10}$  and  $c_{20}$ ), but they are inactivated, or degraded, at some speed proportional to their concentration (terms  $b_{11}$  and  $b_{22}$ ). However, suppose protein  $x_1$  is activated by binding to  $x_2$ ; this interaction in turn inactivates  $x_2$ ; this pathway is modeled by terms  $a_{12}(x_1)$  and  $b_{21}(x_2)$ , which we assume are sigmoidal functions asymptotically constant, consistently with a cooperative, Hill function-type protein interaction.

We can rewrite the above equations as:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -b_{11} & \bar{a}_{12} + \delta_{12} \\ -\bar{b}_{21} - \delta_{21} & -b_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} c_{10} \\ c_{20} \end{bmatrix},$$

where  $\delta_{12} = a_{12}(x_1) - \bar{a}_{12}$  and  $\delta_{21} = a_{21}(x_1) - \bar{a}_{21}$  and where  $\bar{a}_{12} = a_{12}(\infty)$  and  $\bar{b}_{21} = b_{21}(\infty)$ .

If the region near the origin is delimited by a "radius"  $\nu$  sufficiently large, the bounds on  $\delta_{12}$  and  $\delta_{21}$  can be taken arbitrarily tight.

So inside  $N_\nu$ , for large  $\nu > 0$ , we may assume  $|\delta_{12}| \leq \epsilon$  and  $|\delta_{21}| \leq \epsilon$  with small  $\epsilon$ . Since the nominal system, for  $\delta_{12} = \delta_{21} = 0$  is quadratically stable, it admits a quadratic Lyapunov function, inside  $N_\nu$ , this is a Lyapunov function. Inside  $N_\nu$  this is a Lyapunov function for the system because the contribution of terms  $\delta_{12}x_2$  and  $\delta_{21}x_1$  is negligible.

This technique allows us to prove boundedness, but not stability of the original system. Boundedness does imply the existence of equilibria, but their stability may be or may not be verified.

### 2.5.2 Structural Stability of Equilibria

If we can establish boundedness of a system, the existence of equilibria is automatically assured. Then, we can ask two main questions:

- How many equilibria are present?
- Which equilibria are stable?

Several results from the so-called degree theory help us find answers; see, for instance, [31–34]. Here, we recall one particularly useful theorem:

**Theorem 3** *Assume that all the system's equilibria  $\bar{x}^{(i)}$  are strictly positive, and assume that none of them is degenerate, i.e. the Jacobian evaluated at each equilibrium has non-zero determinant. Then:*

$$\sum_i \text{sign det} \left[ -J \left( \bar{x}^{(i)} \right) \right] = 1$$

How does this theorem help us answer our questions? We describe informally three cases that we can immediately discriminate as a consequence of this theorem. Suppose analytical expressions for the Jacobian are available, as a function of a generic equilibrium point.

1. If we can establish that the determinant of  $-J$  is always positive, regardless of specific values for parameters or equilibria, then there is a unique equilibrium.
2. If at an equilibrium point we have  $\text{det}[-J] < 0$ , then such equilibrium must be unstable (because the characteristic polynomial has a negative constant term  $p_0 = \text{det}[-J]$ .) A consequence of Theorem 3 is that other equilibria must exist; if they are not degenerate, then there must be at least two equilibria.
3. If there are two stable equilibria, then necessarily another unstable stable equilibrium must exist.

In a qualitative/parameter-free context, general statements about stability of equilibria are difficult to demonstrate. If we restrict our attention to specific classes of systems, however, we can find structural stability results. We mention a few, well known examples:

- Chemical reaction networks modeled with mass action kinetics: the zero-deficiency theorem [11] guarantees uniqueness of the equilibrium and asymptotic stability of networks satisfying specific structural conditions that do not depend on the reaction rate parameters.

- Monotone systems: if a system is monotone [35], then its Jacobian has nonnegative non-diagonal entries, in other words it is a Metzler matrix. For a Metzler matrix, stability is equivalent to having a characteristic polynomial with all positive coefficients. This property is easy to check analytically in systems of small dimension.
- Planar systems. Plenty of straightforward methods are available to find structural stability conditions.

We conclude this section with a paradox:

**Difficulty:** Structural stability investigation is, generally speaking, an unsolved problem which typically requires a case-by-case study.

**Interest:** Stability is generally of little interest to biologists, because many natural behaviors in biology are known to be (obviously) stable. In other words, formal proofs of stability are not very informative. However, lack of stability of an equilibrium can be a hallmark for other interesting behaviors, such as multistationarity and periodicity.

## 2.6 Conclusions

A property is structurally robust if it is satisfied by a class of models regardless of the specific expressions adopted or of the parameter values in the model. This chapter highlights that qualitative, parameter-free models of molecular networks can be formulated by making general assumptions on the sign, trend and boundedness of the species interactions. Linearization, Lyapunov methods, invariant sets and graphical tests are examples of classical control theoretic tools that can be successfully employed to analyze such qualitative models, often reaching strong conclusions on their admissible dynamic behavior.

Robustness is often tested through simulations, at the price of exhaustive campaigns of numerical trials and, more importantly, with no theoretical guarantee of robustness. We are far from claiming that numerical simulations are useless: they are useful, for instance, to falsify “robustness conjectures” by finding suitable numerical counterexamples. In addition, for very complex systems in which analytical tools cannot be employed, simulations are the only viable method for analysis. A limit of our qualitative modeling and analysis approach is its lack of systematic scalability to complex models. However, the techniques we employed can be successfully used to study a large class of low dimension systems, and are an important complementary tool to simulations and experiments.



## References

1. Ma W, Trusina A, El-Samad H, Lim WA, Tang C (2009) Defining network topologies that can achieve biochemical adaptation. *Cell* 138(4):760–773
2. Prill RJ, Iglesias PA, Levchenko A (2005) Dynamic properties of network motifs contribute to biological network organization. *Public Libr Sci Biol* 3(11):e343
3. Kwon YK, Cho KH (2008) Quantitative analysis of robustness and fragility in biological networks based on feedback dynamics. *Bioinformatics* 24(7):987–994
4. Gómez-Gardenes J, Floría LM (2005) On the robustness of complex heterogeneous gene expression networks. *Biophys Chem* 115:225–229
5. Gorban A, Radulescu O (2007) Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Syst Biol* 1(4):238–246
6. Kartal O, Ebenhöf O (2009) Ground state robustness as an evolutionary design principle in signaling networks. *Public Libr Sci One* 4(12):e8001
7. Aldana M, Cluzel P (2003) A natural class of robust networks. *Proc Nat Acad Sci USA* 100(15):8710–8714
8. Tian T (2004) Robustness of mathematical models for biological systems. *ANZIAM J* 45:C565–C577
9. Shinar G, Milo R, Martínez MR, Alon U (2007) Input–output robustness in simple bacterial signaling systems. *Proc Nat Acad Sci USA* 104:19931–19935
10. Shinar G, Feinberg M (2010) Structural sources of robustness in biochemical reaction networks. *Science* 327(5971):1389–1391
11. Feinberg M (1987) Chemical reaction network structure and the stability of complex isothermal reactors I. The deficiency zero and deficiency one theorems. *Chem Eng Sci* 42:2229–2268
12. Sontag E (2007) Monotone and near-monotone biochemical networks. *Syst Synth Biol* 1:59–87
13. Angeli D, Ferrell JE, Sontag ED (2004) Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc Nat Acad Sci USA* 101(7):1822–1827
14. Jacquez J, Simon C (1993) Qualitative theory of compartmental systems. *Soc Ind Appl Math Rev* 35:43–79
15. Nikolov S, Yankulova E, Wolkenhauer O, Petrov V (2007) Principal difference between stability and structural stability (robustness) as used in systems biology. *Nonlinear Dyn, Psychol, Life Sci* 11(4):413–433
16. Blanchini F, Franco E (2011) Structurally robust biological networks. *Bio Med Central Syst Biol* 5:74
17. Abate A, Tiwari A, Sastry S (2007) Box Invariance for biologically-inspired dynamical systems. In: *Proceedings of the IEEE conference on decision and control*, pp 5162–5167
18. El-Samad H, Prajna S, Papachristodoulou A, Doyle J, Khammash M (2006) Advanced methods and algorithms for biological networks analysis. *Proc IEEE* 94(4):832–853
19. De Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9:67–103
20. Alon U (2006) *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC, UK, USA
21. Kitano H (2002) Systems biology: a brief overview. *Science* 295(5560):1662–1664
22. Goldbeter A, Gérard C, Gonze D, Leloup JC, Dupont G (2012) Systems biology of cellular rhythms. *FEBS Lett* 586:2955–2965
23. Ackermann J, Wlotzka B, McCaskill JS (1998) In vitro DNA-based predator-prey system with oscillatory kinetics. *Bull Math Biol* 60(2):329–354
24. Balagadde FK, Song H, Ozaki J, Collins CH, Barnet M, Arnold FH, Quake SR, You L (2008) A synthetic *Escherichia coli* predator-prey ecosystem. *Mol Syst Biol* 4. <http://dx.doi.org/10.1038/msb.2008.24>
25. Yi TM, Huang Y, Simon MI, Doyle J (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Nat Acad Sci USA* 97(9):4649–4653

26. Drengstig T, Ueda HR, Ruoff P (2008) Predicting perfect adaptation motifs in reaction kinetic networks. *J Phys Chem B* 112(51):16752–16758
27. Santos SDM, Verwee PJ, Bastiaens PIH (2007) Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat Cell Biol* 9(3):324–330
28. Franco E, Blanchini F (2012) Structural properties of the MAPK pathway topologies in PC12 cells. *J Math Biol*
29. Vilar JMG, Guet C, Leibler S (2003) Modeling network dynamics: the lac operon, a case study. *J Cell Biol* 161(3):471–476
30. Blanchini F, Miani S (2008) Set-theoretic methods in control, Vol 22 of *Systems and Control: Foundations and Applications*. Birkhäuser, Boston
31. Ortega R, Campos J (1995) Some applications of the topological degree to stability theory. In: *Topological methods in differential equations and inclusions*. Kluwer Academic Publishing, Dordrecht, pp 377–409
32. Hofbauer J (1990) An index theorem for dissipative semiflows. *Rocky Mt J Math* 20(4):1017–1031
33. Meiss J (2007) *Differential dynamical systems*. SIAM
34. Ambrosetti A, Prodi G (1995) *A primer of nonlinear analysis*. SIAM
35. Smith HL (2008) *Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems*. Am Math Soc

# Chapter 3

## Guaranteeing Spatial Uniformity in Reaction-Diffusion Systems Using Weighted $L^2$ Norm Contractions

Zahra Aminzare, Yusef Shafi, Murat Arcak and Eduardo D. Sontag

**Abstract** We present conditions that guarantee spatial uniformity of the solutions of reaction-diffusion partial differential equations. These equations are of central importance to several diverse application fields concerned with pattern formation. The conditions make use of the Jacobian matrix and Neumann eigenvalues of elliptic operators on the given spatial domain. We present analogous conditions that apply to the solutions of diffusively-coupled networks of ordinary differential equations. We derive numerical tests making use of linear matrix inequalities that are useful in certifying these conditions. We discuss examples relevant to enzymatic cell signaling and biological oscillators. From a systems biology perspective, the paper's main contributions are unified verifiable relaxed conditions that guarantee spatial uniformity of biological processes.

**Keywords** Reaction-diffusion systems · Turing phenomenon · Diffusive instabilities · Compartmental systems · Contraction methods for stability · Matrix measures

---

The authors Zahra Aminzare and Yusef Shafi contributed equally.

---

Z. Aminzare · E. D. Sontag (✉)  
Department of Mathematics, Rutgers University, Piscataway, NJ, USA  
e-mail: sontag@math.rutgers.edu

Z. Aminzare  
e-mail: aminzare@math.rutgers.edu

Y. Shafi · M. Arcak  
Department of Electrical Engineering and Computer Sciences, University of California,  
Berkeley, CA, USA  
e-mail: yusef@eecs.berkeley.edu

M. Arcak  
e-mail: arcak@eecs.berkeley.edu

### 3.1 Introduction

This paper studies reaction-diffusion partial differential equations (PDEs) of the form

$$\frac{\partial u}{\partial t}(\omega, t) = F(u(\omega, t), t) + \mathcal{L}u(\omega, t), \quad (3.1)$$

where  $\mathcal{L}$  denotes a diffusion operator. We prove a two-part result that addresses the question of how the stability of solutions of the PDE relates to stability of solutions of the underlying ordinary differential equation (ODE)  $\frac{dx}{dt}(t) = F(x(t), t)$ . The study of this question is central to many application fields concerned with pattern formation, ranging from biology (morphogenesis developmental biology, species competition and cooperation in ecology, epidemiology) [8, 9, 23] and enzymatic reactions in chemical engineering [24] to spatio-temporal dynamics in semiconductors [21].

The first part of our result shows that when solutions of the ODE have a certain contraction property, namely  $\mu_{2,Q}(J_F(u, t)) < 0$  uniformly on  $u$  and  $t$ , where  $\mu_{2,Q}$  is a logarithmic norm (matrix measure) associated to a  $Q$ -weighted  $L^2$  norm, the associated PDE, subject to no-flux (Neumann) boundary conditions, enjoys a similar property. This result complements a similar result shown in [1] which, while allowing norms  $L^p$  with  $p$  not necessarily equal to 2, had the restriction that it only applied to diagonal matrices  $Q$  and  $\mathcal{L}$  was the standard Laplacian. Logarithmic norm or “contraction” approaches arose in the dynamical systems literature [12, 15, 17], and were extended and much further developed in work by Slotine e.g. [16]; see also [18] for historical comments.

The second, and complementary, part of our result shows that when  $\mu_{2,Q}(J_f(u, t) - \Lambda_2) < 0$ , where  $\Lambda_2$  is a nonnegative diagonal matrix whose entries are the second smallest Neumann eigenvalues of the diffusion operators in (1), the solutions become spatially homogeneous as  $t \rightarrow \infty$ . This result generalizes the previous work [3] to allow for spatially-varying diffusion, and makes a contraction principle implicitly used in [3] explicit.

We next turn to compartmental ordinary differential equations (ODEs), where each compartment represents a well-mixed spatial domain wherein corresponding components in adjacent compartments are coupled by diffusion [11], and present spatial uniformity conditions analogous to those derived for the PDE case. We then derive convex linear matrix inequality [4] tests as in [3] that can be used to certify the conditions. Our discussion is punctuated by several examples of biological interest.

## 3.2 Spatial Uniformity in Reaction-Diffusion PDEs

In this section, we study the reaction-diffusion PDE (3.1), subject to a Neumann boundary condition:

$$\nabla u_i \cdot \mathbf{n}(\xi, t) = 0 \quad \forall \xi \in \partial\Omega, \quad \forall t \in [0, \infty). \quad (3.2)$$

**Assumption 1** In (3.1)–(3.2) we assume:

- $\Omega$  is a bounded domain in  $\mathbb{R}^m$  with smooth boundary  $\partial\Omega$  and outward normal  $\mathbf{n}$ .
- $F: V \times [0, \infty) \rightarrow \mathbb{R}^n$  is a (globally) Lipschitz and twice continuously differentiable vector field with respect to  $x$ , and continuous with respect to  $t$ , with components  $F_i$ :

$$F(x, t) = (F_1(x, t), \dots, F_n(x, t))^T$$

for some functions  $F_i: V \times [0, \infty) \rightarrow \mathbb{R}$ , where  $V$  is a convex subset of  $\mathbb{R}^n$ .

- 

$$\mathcal{L} = \text{diag}(\mathcal{L}_1, \dots, \mathcal{L}_n), \quad \text{and} \quad \mathcal{L}u = (\mathcal{L}_1 u_1, \dots, \mathcal{L}_n u_n)^T,$$

where for each  $i = 1, \dots, n$ ,

$$(\mathcal{L}_i u_i)(\omega, t) = \nabla \cdot (A_i(\omega) \nabla u_i(\omega, t)), \quad (3.3)$$

and  $A_i: \Omega \rightarrow \mathbb{R}^{m \times m}$  is symmetric and there exist  $\alpha_i, \beta_i > 0$  such that for all  $\omega \in \Omega$  and  $\zeta = (\zeta_1, \dots, \zeta_m)^T \in \mathbb{R}^m$ ,

$$\alpha_i |\zeta|^2 \leq \zeta^T A_i(\omega) \zeta \leq \beta_i |\zeta|^2. \quad (3.4)$$

Suppose that  $\mathcal{L}$  has  $r \leq n$  distinct elements  $\mathbf{L}_1, \dots, \mathbf{L}_r$  (up to a scalar). Namely,

$$\text{diag}(\mathcal{L}_1, \dots, \mathcal{L}_{n_1}, \dots, \mathcal{L}_{n-n_r+1}, \dots, \mathcal{L}_n) = \text{diag}(d_{11}, \dots, d_{1n_1}, \dots, d_{r1}, \dots, d_{rn_r}) \text{diag}(\mathbf{L}_1, \dots, \mathbf{L}_1, \dots, \mathbf{L}_r, \dots, \mathbf{L}_r),$$

where  $n_1 + \dots + n_r = n$ . For each  $i = 1, \dots, r$ , let  $D_i$  be an  $n \times n$  diagonal matrix with entries  $[D_i]_{n_{i-1}+j, n_{i-1}+j} = d_{ij}$ , for  $j = 1, \dots, n_i$ ,  $n_0 = 0$  elsewhere. Also for each  $i = 1, \dots, r$ , let  $\mathfrak{L}_i$  be an  $n \times n$  diagonal matrix with identical entries  $\mathbf{L}_i$ . Then  $\mathcal{L}$  can be written as below,

$$\mathcal{L} = \sum_{i=1}^r D_i \mathfrak{L}_i. \quad (3.5)$$

Some times it is easier to use expression (3.5) for  $\mathcal{L}$  to prove theorems in this paper.

For a fixed  $i \in \{1, \dots, n\}$ , let  $\lambda_i^k$  be the  $k$ th Neumann eigenvalue of the operator  $-\mathcal{L}_i$  as in (3.3) ( $\lambda_i^1 = 0$ ,  $\lambda_i^k > 0$  for  $k > 1$ , and  $\lambda_i^k \rightarrow \infty$  as  $k \rightarrow \infty$ ) and  $e_i^k$  be the corresponding normalized eigenfunction:

$$\begin{aligned}\nabla \cdot \left( A_i(\omega) \nabla e_i^k(\omega) \right) &= -\lambda_i^k e_i^k(\omega), \quad \omega \in \Omega \\ \nabla e_i^k(\xi) \cdot \mathbf{n} &= 0, \quad \xi \in \partial\Omega\end{aligned}\tag{3.6}$$

Also for each  $i = 1, \dots, r$ , let  $\lambda_i^k$  be the  $k$ th Neumann eigenvalue of  $-\mathbf{L}_i$ . Note that

$$\Lambda_k = \sum_{i=1}^r \lambda_i^k D_i, \quad \text{where } \Lambda_k = \text{diag} \left( \lambda_1^k, \dots, \lambda_n^k \right).\tag{3.7}$$

For each  $k \in \{1, 2, \dots\}$ , let  $E_i^k$  be the subspace spanned by the first  $k$ th eigenfunctions:

$$E_i^k = \langle e_i^1, \dots, e_i^k \rangle.$$

Now define the map  $\Pi_{k,i}$  on  $L^2(\Omega)$  as follows:

$$\Pi_{k,i}(v) = v - \pi_{k,i}(v),$$

where  $\pi_{k,i}$  is the orthogonal projection map onto  $E_i^{k-1}$ , and we define  $E_i^0 = 0$ . Namely for any  $v = \sum_{j=1}^{\infty} (v, e_i^j) e_i^j$ ,

$$\begin{aligned}\pi_{k,i}(v) &= \sum_{j=1}^{k-1} (v, e_i^j) e_i^j \quad \text{and} \quad \Pi_{k,i}(v) = \sum_{j=k}^{\infty} (v, e_i^j) e_i^j, \quad \text{for } k > 1, \\ \pi_{1,i}(v) &= 0, \quad \text{and} \quad \Pi_{1,i}(v) = v;\end{aligned}\tag{3.8}$$

where  $(x, y) := \int x^T y$ . Note that for any  $i = 1, \dots, n$ ,

$$\Pi_{2,i}(v) = v - \frac{1}{|\Omega|} \int_{\Omega} v.\tag{3.9}$$

For any  $v = (v_1, \dots, v_n)$ , define  $\Pi_k$  as follows:

$$\Pi_k(v) = v - \pi_k(v) \quad \text{where} \quad \pi_k(v) = (\pi_{k,1}(v_1), \dots, \pi_{k,n}(v_n))^T.$$

Observe that  $\pi_k(v)$  is the orthogonal projection map onto  $E_1^{k-1} \times \dots \times E_n^{k-1}$ .

**Definition 1** By a solution of the PDE

$$\begin{aligned}\frac{\partial u}{\partial t}(\omega, t) &= F(u(\omega, t), t) + \mathcal{L}u(\omega, t), \\ \nabla u_i \cdot \mathbf{n}(\xi, t) &= 0 \quad \forall \xi \in \partial\Omega, \quad \forall t \in [0, \infty)\end{aligned}$$

on an interval  $[0, T)$ , where  $0 < T \leq \infty$ , we mean a function  $u = (u_1, \dots, u_n)^T$ , with  $u: \bar{\Omega} \times [0, T) \rightarrow V$ , such that:

1. for each  $\omega \in \bar{\Omega}$ ,  $u(\omega, \cdot)$  is continuously differentiable;
2. for each  $t \in [0, T)$ ,  $u(\cdot, t)$  is in  $\mathbf{Y}$ , where  $\mathbf{Y}$  is defined as the following set:

$$\left\{ v = (v_1, \dots, v_n)^T : \bar{\Omega} \rightarrow V \mid v_i \in C_{\mathbb{R}}^2(\bar{\Omega}), \frac{\partial v_i}{\partial \mathbf{n}}(\xi) = 0, \forall \xi \in \partial\Omega \quad \forall i \right\},$$

where  $C_{\mathbb{R}}^2(\bar{\Omega})$  is the set of twice continuously differentiable functions  $\bar{\Omega} \rightarrow \mathbb{R}$ .

3. for each  $\omega \in \bar{\Omega}$ , and each  $t \in [0, T)$ ,  $u$  satisfies the above PDE.

Theorems on existence and uniqueness of solutions for PDEs such as (3.1)–(3.2) can be found in standard references, e.g. [5, 22].

For any invertible matrix  $Q$ , and any  $1 \leq p \leq \infty$ , and continuous  $u : \Omega \rightarrow \mathbb{R}^n$ , we denote the weighted  $L_{p,Q}$  norm,  $\|u\|_{p,Q} = \|Qu\|_p$ , where  $(Qu)(\omega) = Qu(\omega)$  and  $\|\cdot\|_p$  indicates the norm in  $L^p(\Omega, \mathbb{R}^n)$ .

**Definition 2** Let  $(X, \|\cdot\|_X)$  be a finite dimensional normed vector space over  $\mathbb{R}$  or  $\mathbb{C}$ . The space  $\mathcal{L}(X, X)$  of linear transformations  $M : X \rightarrow X$  is also a normed vector space with the induced operator norm

$$\|M\|_{X \rightarrow X} = \sup_{\|x\|_X=1} \|Mx\|_X.$$

The logarithmic norm  $\mu_X(\cdot)$  induced by  $\|\cdot\|_X$  is defined as the directional derivative of the matrix norm, that is,

$$\mu_X(M) = \lim_{h \rightarrow 0^+} \frac{1}{h} (\|I + hM\|_{X \rightarrow X} - 1),$$

where  $I$  is the identity operator on  $X$ .

In [1], we proved the following lemma:

**Lemma 1** Consider the PDE system (3.1)–(3.2), with  $\mathcal{L} = D\Delta$ , where  $D = \text{diag}(d_1, \dots, d_n)$ . In addition suppose Assumption 1 holds. For some  $1 \leq p \leq \infty$ , and a positive diagonal matrix  $Q$ , let

$$\mu := \sup_{(x,t) \in V \times [0,\infty)} \mu_{p,Q}(J_F(x,t)).$$

(We are using  $\mu_{p,Q}$  to denote the logarithmic norm associated to the norm  $\|Qv\|_p$  in  $\mathbb{R}^n$ .) Then for any two solutions  $u$  and  $v$  of (3.1)–(3.2), we have

$$\|u(\cdot, t) - v(\cdot, t)\|_{p,Q} \leq e^{\mu t} \|u(\cdot, 0) - v(\cdot, 0)\|_{p,Q}.$$

The first part of the following theorem is a generalization of Lemma 1 to non-diagonal  $P$  for the special case of  $p = 2$ . The second part of the theorem is a generalization of Theorem 1 from [3] to spatially-varying diffusion.

**Theorem 1** Consider the reaction-diffusion system (3.1)–(3.2) and suppose Assumption 1 holds. For  $k = 1, 2$ , let

$$\mu_k := \sup_{(x,t) \in V \times [0, \infty)} \mu_{2,P}(J_F(x,t) - \Lambda_k),$$

for a positive symmetric matrix  $P$  such that for any  $i = 1, \dots, r$ :

$$P^2 D_i + D_i P^2 > 0. \quad (3.10)$$

Then for any two solutions, namely  $u$  and  $v$ , of (3.1)–(3.2), we have:

$$\|u(\cdot, t) - v(\cdot, t)\|_{2,P} \leq e^{\mu_1 t} \|u(\cdot, 0) - v(\cdot, 0)\|_{2,P}. \quad (3.11)$$

In addition

$$\|\Pi_2(u(\cdot, t))\|_{2,P} \leq e^{\mu_2 t} \|\Pi_2(u(\cdot, 0))\|_{2,P}. \quad (3.12)$$

Before proving the main theorem of this section, Theorem 1, we first prove the following:

**Lemma 2** Suppose that  $P$  is a positive definite, symmetric matrix and  $M$  is an arbitrary matrix.

1. If  $\mu_{2,P}(M) = \mu$ , then  $QM + M^T Q \leq 2\mu Q$ , where  $Q = P^2$ .
2. If for some  $Q = Q^T > 0$ ,  $QM + M^T Q \leq 2\mu Q$ , then there exists  $P = P^T > 0$  such that  $P^2 = Q$  and  $\mu_{2,P}(M) \leq \mu$ .

*Proof* First suppose  $\mu_{2,P}(M) = \mu$ . By definition of  $\mu$ :

$$\frac{1}{2} \left( PMP^{-1} + (PMP^{-1})^T \right) \leq \mu I.$$

Since  $P$  is symmetric, so is  $P^{-1}$ , so

$$PMP^{-1} + P^{-1}M^T P \leq 2\mu I.$$

Now multiplying the last inequality by  $P$  on the right and the left, we get:

$$P^2 M + M^T P^2 \leq 2\mu P^2.$$

This proves 1. Now assume that for some  $Q = Q^T > 0$ ,  $QM + M^T Q \leq 2\mu Q$ . Since  $Q > 0$ , there exists  $P > 0$  such that  $P^T P = Q$ ; moreover, because  $Q$  is symmetric, so is  $P$ . Hence we have:

$$P^2 M + M^T P^2 \leq 2\mu P^2.$$



Multiplying the last inequality by  $P^{-1}$  from right and from left, we conclude 2.  $\square$

*Remark 1* Observe that for  $Q > 0$ ,

1.

$$QM + M^T Q \leq \mu Q \Rightarrow QM + M^T Q \leq \beta I,$$

where  $\beta = \mu\lambda$  and  $\lambda$  is the smallest eigenvalue of  $Q$ .

2.

$$QM + M^T Q \leq \beta I \Rightarrow QM + M^T Q \leq \gamma Q,$$

where  $\gamma = \frac{\beta}{\lambda'}$  and  $\lambda'$  is the largest eigenvalue of  $Q$ .

We now recall a result following from the Poincaré principle as in [13], which gives a variational characterization of the eigenvalues of an elliptic operator.

**Lemma 3** Consider an elliptic operator as in (3.3) and let  $v = v(\omega)$  be a function not identically zero in  $L^2(\Omega)$  with derivatives  $\frac{\partial v}{\partial \omega_j} \in L^2(\Omega)$  that satisfies the Neumann boundary condition,  $\nabla v(\omega) \cdot \mathbf{n}(\omega) = 0$ , and for any  $j \in \{1, \dots, k-1\}$ ,  $\int_{\Omega} v e_i^j = 0$ . Then the following inequality holds, for any  $k \geq 1$ :

$$\int_{\Omega} \nabla v \cdot (A_i(\omega) \nabla v) d\omega \geq \lambda_i^k \int_{\Omega} v^2 d\omega. \quad (3.13)$$

**Lemma 4** Suppose  $u \in L^2(\Omega)$  satisfies the Neumann boundary conditions. For any  $k \in \{1, 2, \dots\}$ ,

$$(\Pi_k(u), \mathcal{L}\Pi_k(u)) \leq -(\Pi_k(u), \Lambda_k \Pi_k(u)). \quad (3.14)$$

In addition for  $k = 1, 2$  and any  $n \times n$  symmetric matrix  $Q$  with the following property:

$$QD_i + D_i Q > 0 \quad i = 1, \dots, r, \quad (3.15)$$

we have:

$$(\Pi_k(u), Q\mathcal{L}\Pi_k(u)) \leq -(\Pi_k(u), Q\Lambda_k \Pi_k(u)). \quad (3.16)$$

*Proof* Note that by (3.6), for any  $\xi \in \partial\Omega$ ,

$$\nabla \Pi_{k,i}(u_i(\xi)) \cdot \mathbf{n} = \sum_{j=k}^{\infty} (u_i, e_i^j) \nabla e_i^j(\xi) \cdot \mathbf{n} = 0.$$

Also by the definition of  $\Pi_{k,i}$ , for any  $j = 1, \dots, k-1$ ,

$$\int_{\Omega} \Pi_{k,i}(u_i) e_i^j d\omega = 0.$$

Then by this last equality, Green's identity and Lemma 3 we get:

$$\begin{aligned} & (\Pi_k(u), \mathcal{L}\Pi_k(u)) \\ &= \int_{\Omega} \Pi_k(u)^T (\nabla \cdot (A_1(\omega) \nabla \Pi_{k,1}(u_1)), \dots, \nabla \cdot (A_n(\omega) \nabla \Pi_{k,n}(u_n)))^T d\omega \\ &= \sum_{i=1}^n \int_{\Omega} \Pi_{k,i}(u_i) \nabla \cdot (A_i(\omega) \nabla \Pi_{k,i}(u_i)) d\omega \\ &= \sum_{i=1}^n \int_{\partial\Omega} \Pi_{k,i}(u_i) A_i(\omega) \nabla \Pi_{k,i}(u_i) \cdot \mathbf{n} dS \\ &\quad - \sum_{i=1}^n \int_{\Omega} \nabla \Pi_{k,i}(u_i)^T A_i(\omega) \nabla \Pi_{k,i}(u_i) d\omega \\ &\leq - \sum_{i=1}^n \lambda_i^k \int_{\Omega} \Pi_{k,i}^2(u_i) d\omega \\ &= - (\Pi_k(u), \Lambda_k \Pi_k(u)). \end{aligned}$$

Since for each  $i = 1, \dots, r$ ,  $QD_i + D_iQ > 0$ , there exists positive definite symmetric matrix  $M_i$ , such that  $QD_i + D_iQ = 2M_i^T M_i$ . Note that

$$\begin{aligned} 2 (\Pi_k(u), QD_i \mathcal{L}_i \Pi_k(u)) &= (\Pi_k(u), (QD_i + D_iQ) \mathcal{L}_i \Pi_k(u)) \\ &\quad + (\Pi_k(u), (QD_i - D_iQ) \mathcal{L}_i \Pi_k(u)). \end{aligned}$$

A simple calculation shows that  $(\Pi_k(u), (QD_i - D_iQ) \mathcal{L}_i \Pi_k(u)) = 0$ :

Let  $Y = QD_i$ . Then since  $Q$  and  $D_i$  are symmetric,  $Y^T = D_iQ$ . Also let  $x = \Pi_k(u)$  and  $y = Yx = QD_i \Pi_k(u)$ . By the definition of  $\mathcal{L}_i$ ,  $Y \mathcal{L}_i = \mathcal{L}_i Y$ , hence we need to show:

$$(x, \mathcal{L}_i y) = (y, \mathcal{L}_i x).$$

By the definition of  $\mathcal{L}_i$ , it suffices to show that for any  $j = 1, \dots, n$ :

$$(x_j, \mathbf{L}_i y_j) = (y_j, \mathbf{L}_i x_j).$$

This last equality holds by the definition of  $\mathbf{L}_i$ , the Neumann boundary condition, and Green's identity. Therefore, using (3.14), for  $k = 1, 2$ , we get

$$\begin{aligned}
(\Pi_k(u), QD_i \mathcal{L}_i \Pi_k(u)) &= \frac{1}{2} (\Pi_k(u), (QD_i + D_i Q) \mathcal{L}_i \Pi_k(u)) \\
&= \left( \Pi_k(u), M_i^T M_i \mathcal{L}_i \Pi_k(u) \right) \\
&= (M_i \Pi_k(u), M_i \mathcal{L}_i \Pi_k(u)) \\
&= (M_i \Pi_k(u), \mathcal{L}_i M_i \Pi_k(u)) \\
&= (\Pi_k(M_i u), \mathcal{L}_i \Pi_k(M_i u)) \\
&\leq -\lambda_i^k (\Pi_k(M_i u), \Pi_k(M_i u)) \\
&= -\lambda_i^k (\Pi_k(u), QD_i \Pi_k(u)). \tag{3.17}
\end{aligned}$$

Note that by the definition of  $\mathcal{L}_i$ ,  $M_i \mathcal{L}_i = \mathcal{L}_i M_i$ . By (3.8) and (3.9), for any  $i, j = 1, \dots, n$ ,

$$\Pi_{k,i} = \Pi_{k,j} \quad \text{for } k = 1, 2.$$

Therefore  $M_i \Pi_k(u) = \Pi_k(M_i u)$  and for any  $l$ ,  $\Pi_{k,l}(M_i u)$  is orthogonal to  $e_i^1$ . Hence we can apply the Poincaré principle. Now using (3.5) and (3.17), we get:

$$\begin{aligned}
(\Pi_k(u), Q\mathcal{L}\Pi_k(u)) &= \sum_{i=1}^r (\Pi_k(u), QD_i \mathcal{L}_i \Pi_k(u)) \\
&\leq -\sum_{i=1}^r \lambda_i^k (\Pi_k(u), QD_i \Pi_k(u)) \\
&= -(\Pi_k(u), Q\Lambda_k \Pi_k(u)). \tag{3.18}
\end{aligned}$$

The last equality holds by Eq. (3.7).  $\square$

**Lemma 5** Suppose  $u \in L^2(\Omega)$  satisfies the Neumann boundary conditions. For any  $k \in \{1, 2, \dots\}$ ,

$$\Pi_k(\mathcal{L}u) = \mathcal{L}\Pi_k(u).$$

*Proof* By the definition of  $\Pi_k$  and  $\mathcal{L}$ , it is enough to show that for a fixed  $i$  ( $i = 1, \dots, n$ ),

$$\Pi_{k,i}(\mathcal{L}_i u_i) = \mathcal{L}_i \Pi_{k,i}(u_i). \tag{3.19}$$

Using the fact that  $\mathcal{L}_i e_i^j = -\lambda_i^j e_i^j$ , the right hand side of (3.19) becomes:

$$\mathcal{L}_i \Pi_{k,i}(u_i) = \mathcal{L}_i \sum_{i=k}^{\infty} (u_i, e_i^j) e_i^j = \sum_{i=k}^{\infty} (u_i, e_i^j) \mathcal{L}_i e_i^j = -\sum_{i=k}^{\infty} (u_i, e_i^j) \lambda_i^j e_i^j;$$

and using the orthogonality of the  $e_i^j$ 's, the left hand side of (3.19) becomes:

$$\begin{aligned}
\Pi_{k,i}(\mathcal{L}_i u_i) &= \sum_{j=k}^{\infty} (\mathcal{L}_i u_i, e_i^j) e_i^j = \sum_{j=k}^{\infty} \left( \mathcal{L}_i \sum_{l=1}^{\infty} (u_i, e_i^l) e_i^l, e_i^j \right) e_i^j \\
&= \sum_{j=k}^{\infty} \left( \sum_{l=1}^{\infty} (u_i, e_i^l) \mathcal{L}_i e_i^l, e_i^j \right) e_i^j \\
&= - \sum_{j=k}^{\infty} \left( \sum_{l=1}^{\infty} (u_i, e_i^l) \lambda_i^l e_i^l, e_i^j \right) e_i^j \\
&= - \sum_{j=k}^{\infty} (u_i, e_i^j) \lambda_i^j e_i^j.
\end{aligned}$$

Hence (3.19) holds.  $\square$

**Lemma 6** *Let  $w = u - x$ , where  $u$  is a solution of (3.1)–(3.2) and  $x = \pi_2(u)$  or  $x = v$  is another solution of (3.1)–(3.2). Note that for  $x = v$ ,  $w = \Pi_1(u - v)$  and for  $x = \pi_2(u)$ ,  $w = \Pi_2(u)$ . For a positive, symmetric matrix  $Q$ , let*

$$\Phi(w) := \frac{1}{2}(w, Qw).$$

Then

$$\frac{d\Phi}{dt}(w) = (w, Q(F(u, t) - F(x, t))) + (w, Q\mathcal{L}w). \quad (3.20)$$

*Proof* For  $x = v$ ,

$$\begin{aligned}
\frac{d\Phi}{dt}(w) &= (u - v, Q \frac{d}{dt}(u - v)) \\
&= (w, Q(F(u, t) - F(v, t))) + (w, Q\mathcal{L}(u - v)) \\
&= (w, Q(F(u, t) - F(x, t))) + (w, Q\mathcal{L}w).
\end{aligned}$$

For  $x = \pi_2(u)$ , i.e.  $w = \Pi_2(u)$ ,

$$\begin{aligned}
\frac{d\Phi}{dt}(w) &= (\Pi_2(u), Q \frac{d}{dt}(\Pi_2(u))) \\
&= (\Pi_2(u), Q\Pi_2(F(u, t))) + (w, Q\Pi_2(\mathcal{L}u)) \\
&= (\Pi_2(u), Q\Pi_2(F(u, t))) + (w, Q\mathcal{L}\Pi_2(u)) \quad \text{by Lemma 5} \\
&= (\Pi_2(u), Q(F(u, t) - \pi_2(F(u, t)))) + (w, Q\mathcal{L}w) \\
&= (\Pi_2(u), Q(F(u, t) - F(\pi_2(u), t))) + (w, Q\mathcal{L}w) \\
&\quad + (\Pi_2(u), Q(\pi_2(F(u, t)) - F(\pi_2(u), t))) \\
&= (w, Q(F(u, t) - F(x, t))) + (w, Q\mathcal{L}w).
\end{aligned}$$

Note that the last equality holds because  $Q(\pi_2(F(u, t)) - F(\pi_2(u), t))$  is independent of  $\omega$  and  $\int_{\Omega} \Pi_{2,i}(u) = 0$ .  $\square$

Now we are ready to prove Theorem 1.

**Proof of Theorem 1**

*Proof* By Lemma 2,

$$Q(J_F - \Lambda_k) + (J_F - \Lambda_k)^T Q \leq 2\mu_k Q, \quad (3.21)$$

where  $Q = P^2$ . Define  $w$  and  $\Phi(w)$  as in Lemma 6 for  $Q = P^2$ . Since  $\Phi(w) = \frac{1}{2} \|Pw\|_2^2$ , to prove (3.11) and (3.12), it's enough to show that for  $k = 1, 2$

$$\frac{d}{dt} \Phi(w) \leq 2\mu_k \Phi(w).$$

Note that by Lemma 4, and the fact that  $w = \Pi_1(u - v)$  or  $w = \Pi_2(u)$ , the second term of the right hand side of (3.20),  $\frac{d}{dt} \Phi(w)$ , satisfies:

$$(w, Q\mathcal{L}w) \leq -(w, Q\Lambda_k w). \quad (3.22)$$

Next, by the Mean Value Theorem for integrals, and using (3.21), we rewrite the first term of the right hand side of (3.20) as follows:

$$\begin{aligned} (w, Q(F(u, t) - F(x, t))) &= \int_{\Omega} w^T(\omega, t) Q(F(u(\omega, t), t) - F(x, t)) d\omega \\ &= \int_{\Omega} w^T(\omega, t) Q \int_0^1 J_F(x + sw(\omega, t), t) \cdot w(\omega, t) ds d\omega \\ &= \int_0^1 \int_{\Omega} w^T(\omega, t) Q J_F(x + sw(\omega, t), t) \cdot w(\omega, t) d\omega ds. \end{aligned}$$

This last equality together with (3.22) imply:

$$\begin{aligned} &(w, Q(F(u, t) - F(x, t))) + (w, Q\mathcal{L}w) \\ &\leq \int_0^1 \int_{\Omega} w^T(\omega, t) Q (J_F(x + sw(\omega, t), t) - \Lambda_k) \\ &\quad \cdot w(\omega, t) d\omega ds \\ &\leq \frac{2\mu_k}{2} \int_0^1 ds \int_{\Omega} w^T Q w d\omega \end{aligned}$$

$$\begin{aligned}
 &= \frac{2\mu_k}{2} \int_{\Omega} w^T Q w \, d\omega \\
 &= 2\mu_k \Phi(w).
 \end{aligned}$$

Therefore

$$\frac{d\Phi}{dt}(w) \leq 2\mu_k \Phi(w).$$

This last inequality implies (3.11) and (3.12) for  $k = 1$  and  $k = 2$  respectively.  $\square$

**Corollary 1** *In Theorem 1, if  $\mu_1 < 0$ , then (3.1)–(3.2) is contracting, meaning that solutions converge (exponentially) to each other, as  $t \rightarrow +\infty$  in the weighted  $L_{2,P}$  norm:*

$$\|u(\cdot, t) - v(\cdot, t)\|_{2,P} \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

**Corollary 2** *In Theorem 1, if  $\mu_2 < 0$ , then solutions converge (exponentially) to uniform solutions, as  $t \rightarrow +\infty$  in the weighted  $L_{2,P}$  norm:*

$$\|\Pi_2(u(\cdot, t))\|_{2,P} \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

Note that (3.16) doesn't necessarily hold for any  $k > 2$ , since for  $k > 2$ , the  $\Pi_{k,i}$ 's could be different for different  $i$ 's. In the following lemma we provide a condition for which (3.16) holds for any  $k$ .

**Lemma 7** *Assume  $P\mathcal{L} = \mathcal{L}P$ , where  $P$  is a positive, symmetric  $n \times n$  matrix and  $P^2 = Q$ . Then for any  $k = 1, 2, \dots$*

$$(\Pi_k(u), Q\mathcal{L}\Pi_k(u)) \leq -(\Pi_k(u), Q\Lambda_k\Pi_k(u)).$$

*Proof* The proof is analogous to the proof of (3.16), using the fact that  $P\mathcal{L} = \mathcal{L}P$  implies that  $P$  is diagonal (if all  $\mathcal{L}_i$ 's are different) or block diagonal (for equal Laplacian operators).  $\square$

**Remark 2** Note that Theorem 1 is valid if  $P\mathcal{L} = \mathcal{L}P$  is assumed instead of (3.15), because (3.16) holds by Lemma 7 and this is all that is needed in the proof. In the following theorem we use this condition to generalize the result of Theorem 1 for any arbitrary  $k$  but restricted to linear systems. We omit the proof, which is analogous.

**Theorem 2** *Consider the reaction-diffusion system (3.1)–(3.2) and suppose Assumption 1 holds. In addition assume that  $F$  is a linear function. For  $k \in \{1, 2, \dots\}$ , let*

$$\mu_k := \sup_{(x,t) \in V \times [0,\infty)} \mu_{2,P}(J_F(x, t) - \Lambda_k),$$

*for a positive symmetric matrix  $P$  such that  $P\mathcal{L} = \mathcal{L}P$ . Then for any two solutions, namely  $u$  and  $v$ , of (3.1)–(3.2), we have:*

$$\|\Pi_k(u(\cdot, t) - v(\cdot, t))\|_{2,P} \leq e^{\mu_k t} \|\Pi_k(u(\cdot, 0) - v(\cdot, 0))\|_{2,P}. \quad (3.23)$$

*Example 1* In [1] we studied the following system:

$$\begin{aligned} x_t &= z - \delta x + k_1 y - k_2(S_Y - y)x + d_1 \Delta x \\ y_t &= -k_1 y + k_2(S_Y - y)x + d_2 \Delta y, \end{aligned}$$

where  $(x(t), y(t)) \in V = [0, \infty) \times [0, S_Y]$  for all  $t \geq 0$  ( $V$  is convex), and  $S_Y, k_1, k_2, \delta, d_1$ , and  $d_2$  are arbitrary positive constants.

This two-dimensional system is a prototype for a large class of models of enzymatic cell signaling as well as transcriptional components. Generalizations to systems of higher dimensions, representing networks of such systems, may be studied as well [19].

In [19], it has been shown that for  $p = 1$ , there exists a positive, diagonal matrix  $Q$ , independent of  $d_1$  and  $d_2$ , such that for all  $(x, y) \in V$ ,  $\mu_{1,Q}(J_F(x, y)) < 0$ ; and then by Lemma 1 one concludes that the system is contractive.

Specifically, [1] showed that for any positive, diagonal matrix  $Q$  and any  $p > 1$ , there exists  $(x, y) \in V$  such that  $\mu_{p,Q}(J_F(x, y)) \geq 0$ , where

$$F = (z - \delta x + k_1 y - k_2(S_Y - y)x, -k_1 y + k_2(S_Y - y)x)^T,$$

and

$$J_F = \begin{pmatrix} -\delta - a & b \\ a & -b \end{pmatrix},$$

with  $a = k_2(S_Y - y) \in [0, k_2 S_Y]$  and  $b = k_1 + k_2 x \in [k_1, \infty)$ .

Now we show that there exists some positive, symmetric (but non-diagonal) matrix  $P$  such that for all  $(x, y) \in V$ ,  $\mu_{2,P} J_F(x, y) < 0$  and  $P^2 D + D P^2 > 0$ , where  $D = \text{diag}(d_1, d_2)$ . Then by Theorem 1 (for  $r = 1$  and  $\mathbf{L}_i u_i = \Delta u_i$ ), and Corollary 1, one can conclude that the system is contractive.

**Claim** Let  $Q = \begin{bmatrix} 1 & 1 \\ 1 & q \end{bmatrix}$ , where  $q > \max \left\{ 1 + \frac{\delta}{4k_1}, \left( \frac{1}{2\sqrt{d}} + \frac{\sqrt{d}}{2} \right)^2 \right\}$ , and  $d = \frac{d_1}{d_2}$ . Then  $Q J_F + (Q J_F)^T < 0$  and  $Q D + D Q > 0$ .

Note that  $Q$  is symmetric and positive (because  $q > 1$ ).

*Proof of Claim* We first compute  $Q J_F$ :

$$\begin{bmatrix} 1 & 1 \\ 1 & q \end{bmatrix} \begin{bmatrix} -\delta - a & b \\ a & -b \end{bmatrix} = \begin{bmatrix} -\delta & 0 \\ -\delta + (q-1)a & -b(q-1) \end{bmatrix}.$$

So

$$Q J_F + (J_F Q)^T = \begin{bmatrix} -2\delta & -\delta + (q-1)a \\ -\delta + (q-1)a & -2b(q-1) \end{bmatrix}.$$

To show  $QJ_F + J_F^T Q < 0$ , we show that  $\det(QJ_F(x, y) + J_F^T(x, y)Q) > 0$  for all  $(x, y) \in V$ :

$$\det(QJ_F + J_F^T Q) = 4\delta b(q - 1) - (-\delta + (q - 1)a)^2.$$

Note that for any  $q > 1$ ,  $f(a) := (-\delta + (q - 1)a)^2 \leq \delta^2$  on  $[0, k_2 S_Y]$ , and  $g(b) := 4\delta b(q - 1) \geq 4\delta k_1(q - 1)$  on  $[k_1, \infty]$ . So to have  $\det > 0$ , it's enough to have  $4\delta k_1(q - 1) - \delta^2 > 0$ , i.e.  $q - 1 > \frac{\delta^2}{4\delta k_1}$ , i.e.  $q > 1 + \frac{\delta}{4k_1}$ . Now we compute  $QD + DQ$ :

$$QD + DQ = \begin{bmatrix} 2d_1 & d_1 + d_2 \\ d_1 + d_2 & 2qd_2 \end{bmatrix}.$$

$QD + DQ > 0$  if and only if  $\det(QD + DQ) > 0$ , i.e.  $4d_1 d_2 q - (d_1 + d_2)^2 > 0$ , i.e.  $q > \left(\frac{1}{2\sqrt{d}} + \frac{\sqrt{d}}{2}\right)^2$ , where  $d = \frac{d_1}{d_2}$ .  $\square$

Now by Remark 1 and Lemma 2, for  $P = \sqrt{Q}$ ,  $\mu_{2,P}(J_F(x, y)) < 0$ , for all  $(x, y) \in V$ .

*Example 2* We now provide an example of a class of reaction-diffusion systems  $x_t = F(x) + D\Delta x$ , with  $x \in V$  ( $V$  convex), which satisfy the following conditions:

1. For some positive definite, diagonal matrix  $Q$ ,  $\sup_{x \in V} \mu_{1,Q}(J_F(x)) < 0$  (and hence by Lemma 1, these systems are contractive).
2. For any positive definite, symmetric (not necessarily diagonal) matrix  $P$ ,  $\sup_{x \in V} \mu_{2,P}(J_F(x)) \not\leq 0$ .

Consider two variable systems of the following type

$$x_t = -f_1(x) + g_1(y) + d_1 \Delta x \tag{3.24}$$

$$y_t = f_2(x) - g_2(y) + d_2 \Delta y, \tag{3.25}$$

where  $d_1, d_2$  are positive constants and  $(x, y) \in V = [0, \infty) \times [0, \infty)$ . The functions  $f_i$  and  $g_i$  take non-negative values. Systems of this form model a case where  $x$  decays according to  $f_1$ ,  $y$  decays according to  $g_2$ , and there is a positive feedback from  $y$  to  $x$  ( $g_1$ ) and a positive feedback from  $x$  to  $y$  ( $f_2$ ).

**Lemma 8** In system (3.24)–(3.25), let  $J$  be the Jacobian matrix of

$$(-f_1(x) + g_1(y), f_2(x) - g_2(y))^T.$$

In addition, assume that the following conditions hold for some  $\lambda > 0$ , and  $\mu > 0$  and all  $(x, y) \in V$ :



1.  $-f'_1(x) + \lambda|f'_2(x)| < -\mu < 0$ ;
2.  $-g'_2(y) + \frac{1}{\lambda}|g'_1(y)| < -\mu < 0$ ;
3. for any  $p_0 \in \mathbb{R}$

$$\lim_{y \rightarrow \infty} \frac{(g'_1(y) - p_0 g'_2(y))^2}{g'_2(y)} = \infty.$$

Then

1. for every  $(x, y) \in V$ ,  $\mu_{1,Q}(J(x, y)) < 0$ , where  $Q = \text{diag}(1, \lambda)$ ; and
2. for each positive definite, symmetric matrix  $P$ , there exists some  $(x, y) \in V$ , such that  $\mu_{2,P}(J(x, y)) \geq 0$ .

*Proof* The proof of  $\mu_{1,Q}(J(x, y)) < 0$  is straightforward from the definition of  $\mu_{1,Q}$  and conditions 1 and 2. Now we show that for any positive matrix  $P = \begin{bmatrix} p_1 & p \\ p & p_2 \end{bmatrix}$ , there exists some  $(x_0, y_0) \in V$  such that  $\mu_{2,P}(J(x_0, y_0)) \geq 0$ . By Lemma 2, it's enough to show that for some  $(x_0, y_0) \in V$ ,  $PJ(x_0, y_0) + J^T(x_0, y_0)P \not< 0$ . We compute:

$$\begin{aligned} PJ &= \begin{bmatrix} p_1 & p \\ p & p_2 \end{bmatrix} \begin{bmatrix} -f'_1(x) & g'_1(y) \\ f'_2(x) & -g'_2(y) \end{bmatrix} \\ &= \begin{bmatrix} -p_1 f'_1(x) + p f'_2(x) & p_1 g'_1(y) - p g'_2(y) \\ -p f'_1(x) + p_2 f'_2(x) & p g'_1(y) - p_2 g'_2(y) \end{bmatrix}. \end{aligned}$$

Therefore,  $PJ + (PJ)^T$  is equal to

$$\begin{bmatrix} 2(-p_1 f'_1(x) + p f'_2(x)) & p_1 g'_1(y) - p g'_2(y) - p f'_1(x) + p_2 f'_2(x) \\ p_1 g'_1(y) - p g'_2(y) - p f'_1(x) + p_2 f'_2(x) & 2(p g'_1(y) - p_2 g'_2(y)) \end{bmatrix}.$$

(not showing  $x$  and  $y$  arguments in  $f'_1$  and  $f'_2$  for simplicity). Now fix  $x_0 \in [0, \infty)$  and let

$$A := 2(-p_1 f'_1(x_0) + p f'_2(x_0)),$$

and

$$B := -p f'_1(x_0) + p_2 f'_2(x_0).$$

Then  $\det(PJ + (PJ)^T)$  is equal to

$$2A(p g'_1(y) - p_2 g'_2(y)) - (p_1 g'_1(y) - p g'_2(y) + B)^2. \quad (3.26)$$

We will show that  $\det < 0$ . Dividing both sides of (3.26) by  $p_1^2 g'_2(y)$ , we get:

$$\begin{aligned}
\frac{\det(PJ + (PJ)^T)}{p_1^2 g_2'(y)} &= \frac{2A(pg_1'(y) - p_2g_2'(y))}{p_1^2 g_2'(y)} \\
&\quad - \frac{(g_1'(y) - p_0g_2'(y) + B')^2}{g_2'(y)} \\
&= A'p \frac{g_1'(y)}{g_2'(y)} - A'p_2 \\
&\quad - \frac{(g_1'(y) - p_0g_2'(y))^2}{g_2'(y)} - 2B' \frac{g_1'(y)}{g_2'(y)} \\
&\quad + 2B'p_0 - \frac{B'^2}{g_2'(y)}
\end{aligned}$$

where  $p_0 = \frac{p}{p_1}$ ,  $A' = \frac{2A}{p_1^2}$ , and  $B' = \frac{B}{p_1}$ .

(Note that  $p_1^2 g_2'(y) > 0$  because by condition 2,  $g_2' \geq \mu > 0$ , and  $P > 0$  implies  $p_1 \neq 0$ .)

By condition 2,  $0 \leq \frac{g_1'(y)}{g_2'(y)} \leq \lambda < \infty$  for all  $y$ . Now using condition 3, we can find  $y$  large enough such that  $\det < 0$ .

Since  $\det(PJ(x_0, y_0) + (PJ(x_0, y_0))^T) < 0$  for some  $(x_0, y_0) \in V$ , the matrix  $PJ + (PJ)^T$  has one positive eigenvalue. Therefore  $PJ + (PJ)^T \not\leq 0$ .  $\square$

*Example 3* As a concrete example, take the following system

$$\begin{aligned}
x_t &= -x + y^{2+\epsilon} + d_1 \Delta x \\
y_t &= \delta x - (y^3 + y^{2+\epsilon} + dy) + d_2 \Delta y,
\end{aligned}$$

where  $0 < \delta < 1$ ,  $0 < \epsilon \ll 1$ ,  $d, d_1$ , and  $d_2$  are positive constants and  $(x, y) \in V = [0, \infty) \times [0, \infty)$ .

In this example we show that, the system is contractive in a weighted  $L^1$  norm; while for any positive, symmetric matrix  $P$ , and some  $(x, y) \in V$ ,  $\mu_{2,P} J_F(x, y) \not\leq 0$ . To this end, we verify the conditions of Lemma 8.

For any  $(x, y) \in V$ , we take in Lemma 8,  $\lambda = 1$ , and any  $\mu \in (0, \min\{d, 1 - \delta\})$ :

1.  $-1 + \delta < 0$ , because  $0 < \delta < 1$ .
2.  $-(3y^2 + (2 + \epsilon)y^{1+\epsilon} + d) + (2 + \epsilon)y^{1+\epsilon} = -3y^2 - d \leq -d < 0$ .
3. For any  $p_0 \in \mathbb{R}$ ,

$$\lim_{y \rightarrow \infty} \frac{((1 - p_0)(2 + \epsilon)y^{1+\epsilon} - p_0(3y^2 + d))^2}{3y^2 + (2 + \epsilon)y^{1+\epsilon} + d} = \infty$$

So the conditions in Lemma 8 are verified.  $\square$

### 3.3 Spatial Uniformity in Diffusively-Coupled Systems of ODEs

We next consider a compartmental ODE model where each compartment represents a spatial domain interconnected with the other compartments over an undirected graph:

$$\dot{u}(t) = \tilde{F}(u(t)) - \mathcal{L}u(t). \quad (3.27)$$

Recall that if  $A = (a_{ij})$  is an  $m \times n$  matrix and  $B = (b_{ij})$  is a  $p \times q$  matrix, then the Kronecker product, denoted by  $A \otimes B$ , is the  $mp \times nq$  block matrix defined as follows:

$$A \otimes B := \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix},$$

where  $a_{ij}B$  denote the following  $p \times q$  matrix:

$$a_{ij}B := \begin{bmatrix} a_{ij}b_{11} & \dots & a_{ij}b_{1q} \\ \vdots & \ddots & \vdots \\ a_{ij}b_{p1} & \dots & a_{ij}b_{pq} \end{bmatrix}.$$

The following are some properties of Kronecker product:

1.  $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ ;
2.  $(A \otimes B)^T = A^T \otimes B^T$ .
3. Suppose that  $A$  and  $B$  are square matrices of size  $n$  and  $m$  respectively. Let  $\lambda_1, \dots, \lambda_n$  be the eigenvalues of  $A$  and  $\mu_1, \dots, \mu_m$  be those of  $B$  (listed according to multiplicity). Then the eigenvalues of  $A \otimes B$  are  $\lambda_i \mu_j$  for  $i = 1, \dots, n$ , and  $j = 1, \dots, m$ .

**Assumption 2** In (3.27), we assume:

- For a fixed convex subset of  $\mathbb{R}^n$ , say  $V$ ,  $\tilde{F}: V^N \rightarrow \mathbb{R}^{nN}$  is a function of the form:

$$\tilde{F}(u) = \left( F(u^1)^T, \dots, F(u^N)^T \right)^T,$$

where  $u = ((u^1)^T, \dots, (u^N)^T)^T$ , with  $u^i \in V$  for each  $i$ , and  $F: V \rightarrow \mathbb{R}^n$  is a (globally) Lipschitz function.

- For any  $u \in V^N$  we define  $\|u\|_{p,Q}$  as follows:

$$\|u\|_{p,Q} = \left\| \left( \|Qu^1\|_p, \dots, \|Qu^N\|_p \right)^T \right\|_p,$$

where  $Q$  is a symmetric and positive definite matrix and  $1 \leq p \leq \infty$ .

With a slight abuse of notation, we use the same symbol for a norm in  $\mathbb{R}^n$ :

$$\|x\|_{p,Q} := \|Qx\|_p.$$

- $u: [0, \infty) \rightarrow V^N$  is a continuously differentiable function.
- 

$$\mathcal{L} = \sum_{i=1}^n L_i \otimes E_i,$$

where for any  $i = 1, \dots, n$ ,  $L_i \in \mathbb{R}^{N \times N}$  is a symmetric positive semidefinite matrix and  $L\mathbf{1}_N = 0$ , where  $\mathbf{1}_N = (1, \dots, 1)^T \in \mathbb{R}^N$ . The matrix  $L_i$  is the symmetric generalized graph Laplacian (see, e.g., [10]) that describes the interconnections among component subsystems. For any  $i = 1, \dots, n$ ,  $E_i = e_i e_i^T \in \mathbb{R}^{n \times n}$  is the product of the  $i$ th standard basis vector  $e_i$  multiplied by its transpose.

Similar to the PDE case, we assume that there exists  $r \leq n$  distinct matrices,  $\mathbf{L}_1, \dots, \mathbf{L}_r$  such that

$$\begin{aligned} & \text{diag}(L_1, \dots, L_{n_1}, \dots, L_{n-n_r+1}, \dots, L_n) \\ &= \text{diag}(d_{11}, \dots, d_{1n_1}, \dots, d_{r1}, \dots, d_{rn_r}) \text{diag}(\mathbf{L}_1, \dots, \mathbf{L}_1, \dots, \mathbf{L}_r, \dots, \mathbf{L}_r), \end{aligned}$$

where  $n_1 + \dots + n_r = n$ . For each  $i = 1, \dots, r$ , let  $D_i$  be an  $n \times n$  diagonal matrix with entries  $[D_i]_{n_{i-1}+j, n_{i-1}+j} = d_{ij}$ , for  $j = 1, \dots, n_i$ ,  $n_0 = 0$  elsewhere. Therefore we can write  $\mathcal{L}$  as follows:

$$\mathcal{L} = \sum_{i=1}^r \mathbf{L}_i \otimes D_i \tag{3.28}$$

For a fixed  $i \in \{1, \dots, n\}$ , let  $\lambda_i^k$  be the  $k$ th eigenvalue of the matrix  $L_i$  and  $e_i^k$  be the corresponding normalized eigenvector. Also for a fixed  $i \in \{1, \dots, r\}$ , let  $\lambda_i^k$  be the  $k$ th eigenvalue of the matrix  $\mathbf{L}_i$ . Note that

$$\Lambda_k = \sum_{i=1}^r \lambda_i^k D_i, \tag{3.29}$$

where  $\Lambda_k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$ .

For each  $k \in \{1, 2, \dots, N\}$ , let  $E_i^k$  be the subspace spanned by the first  $k$ th eigenvectors:

$$E_i^k = \langle e_i^1, \dots, e_i^k \rangle.$$

Now let  $\pi_{k,i}$  be the orthogonal projection map from  $\mathbb{R}^N$  onto  $E_i^{k-1}$ . Namely for any  $v = \sum_{j=1}^N (v \cdot e_i^j) e_i^j$ ,

$$\pi_{k,i}(v) = \sum_{j=1}^{k-1} (v \cdot e_i^j) e_i^j,$$

for  $1 < k \leq N$  and  $\pi_{1,i}(v) = 0$ .

Now for  $u = (u^1, \dots, u^N)$  with  $u^j \in \mathbb{R}^n$ , define  $\pi_k(u)$  as follows:

$$\pi_k(u) = \sum_{j=1}^n (\pi_{j,k}(u_j))^T \otimes e_j, \quad (3.30)$$

for  $1 < k \leq N$ , where  $u_j := (u^1 \cdot e_j, \dots, u^N \cdot e_j)^T$ ; and  $\pi_1(u) = 0$ .

Note that for each  $k$  and any  $u, v \in \mathbb{R}^{nN}$ ,

$$(u - \pi_k(u))^T \pi_k(v) = \sum_{j=1}^n (u_j - \pi_{j,k}(u_j))^T \pi_{j,k}(v_j) = 0. \quad (3.31)$$

We also can define  $\pi_k(u)$  as follows:

For  $i = 1, \dots, n$ , let  $e^i := \sum_{j=1}^N e_i^j \otimes e_j$ . It is straightforward to show that  $e^1, \dots, e^n$  are linearly independent and for any  $i, j \in \{1, \dots, n\}$ ,  $e^i \cdot e^j = 0$ . Hence one can extend  $\{e^i\}_{1 \leq i \leq n}$  to an orthogonal basis for  $\mathbb{R}^{nN}$ ,  $\{e^i\}_{1 \leq i \leq nN}$ . Then for each  $k = 2, \dots, nN$ , and any  $u \in \mathbb{R}^{nN}$ ,

$$\pi_k(u) = \sum_{j=1}^{k-1} (u \cdot e^j) e^j,$$

and  $\pi_1(u) = 0$ . Note that for  $k = 1, \dots, n$ , this definition is compatible with (3.30).

We now state Courant-Fischer minimax theorem, from [14].

**Lemma 9** *Let  $L$  be a symmetric, positive semidefinite matrix in  $\mathbb{R}^{N \times N}$ . Let  $\lambda^1 \leq \dots \leq \lambda^N$  be  $N$  eigenvalues with  $e^1, \dots, e^N$  corresponding normalized orthogonal eigenvectors. For any  $v \in \mathbb{R}^N$ , if  $v^T e^j = 0$  for  $1 \leq j \leq k-1$ , then*

$$v^T L v \geq \lambda^k v^T v.$$

**Lemma 10** *Let  $w := u - x$ , where  $u$  is a solution of (3.27) and  $x = v$  is another solution of (3.27) or  $x = \pi_2(u)$ , i.e.  $x = \mathbf{1}_N \otimes \left(\frac{1}{N} \sum_{j=1}^N u^j\right)$ . For a positive, symmetric matrix  $Q$ , let*

$$\Phi(w) := \frac{1}{2} w^T (I_N \otimes Q) w.$$

Then

$$\frac{d\Phi}{dt}(w) = w^T (I_N \otimes Q) (\tilde{F}(u, t) - \tilde{F}(x, t)) - w^T (I_N \otimes Q) \mathcal{L}w. \quad (3.32)$$

*Proof* When  $x = v$ , the claim is trivial because both  $u$  and  $v$  satisfy (3.27). When  $x = \pi_2(u)$ , then, by orthogonality, Eq. (3.31), and the definition of  $\pi_2$ , we have:

$$\begin{aligned} \frac{d\Phi}{dt}(w) &= (u - \pi_2(u))^T (I_N \otimes Q) (\tilde{F}(u, t) - \pi_2(\tilde{F}(u, t))) + w^T (I_N \otimes Q) \mathcal{L}w \\ &= (u - \pi_2(u))^T (I_N \otimes Q) \tilde{F}(u, t) + w^T (I_N \otimes Q) \mathcal{L}w \\ &= (u - \pi_2(u))^T (I_N \otimes Q) (\tilde{F}(u, t) - \tilde{F}(\pi_2(u), t)) + w^T (I_N \otimes Q) \mathcal{L}w, \end{aligned}$$

The last equality holds because

$$\begin{aligned} (u - \pi_2(u))^T (I_N \otimes Q) \tilde{F}(\pi_2(u), t) &= \sum_{j=1}^N (u^j - \bar{u}) Q F(\bar{u}) \\ &= \left( \sum_{j=1}^N u^j - N\bar{u} \right) Q F(u) = 0, \end{aligned}$$

where  $\bar{u} = \frac{1}{N} \sum_{j=1}^N u^j$ .

**Theorem 3** Consider the ODE system (3.27) and suppose Assumption 2 holds. For  $k = 1, 2$ , let

$$\mu_k := \sup_{(x,t) \in V \times [0, \infty)} \mu_{2,P}(J_F(x, t) - \Lambda_k),$$

for a positive symmetric matrix  $P$  such that for every  $i = 1, \dots, r$ ,

$$P^2 D_i + D_i P^2 > 0.$$

Then for any two solutions, namely  $u$  and  $v$ , of (3.27), we have:

$$\|(u - v)(t)\|_{2,P} \leq e^{\mu_1 t} \|(u - v)(0)\|_{2,P}. \quad (3.33)$$

In addition

$$\|(u - \pi_2(u))(t)\|_{2,P} \leq e^{\mu_2 t} \|(u - \pi_2(u))(0)\|_{2,P}. \quad (3.34)$$

*Proof* By Lemma 2,

$$Q(J_F - \Lambda_k) + (J_F - \Lambda_k)^T Q \leq 2\mu_k Q, \quad (3.35)$$

where  $Q = P^2$ . Define  $w$  and  $\Phi(w)$  as in Lemma 10 for  $Q = P^2$ . Since  $\Phi(w) = \frac{1}{2} \|Pw\|_2^2$ , to prove (3.33) and (3.34), it's enough to show that for  $k = 1, 2$

$$\frac{d}{dt} \Phi(w) \leq 2\mu_k \Phi(w).$$

We rewrite the second term of the right hand side of (3.32) as follows. Since  $Q = P^2$  and  $P^2 D_i + D_i P^2 > 0$ , there exists symmetric, positive definite matrices  $M_i$  such that  $Q D_i + D_i Q = 2M_i^T M_i$ .

$$\begin{aligned} w^T (I_N \otimes Q) \mathcal{L} w &= w^T (I_N \otimes Q) \left( \sum_{i=1}^r \mathbf{L}_i \otimes D_i \right) w \\ &= w^T \left( \sum_{i=1}^r I_N \mathbf{L}_i \otimes Q D_i \right) w \\ &= \frac{1}{2} \sum_{i=1}^r w^T (\mathbf{L}_i \otimes (Q D_i + D_i Q)) w \\ &= \sum_{i=1}^r w^T (\mathbf{L}_i \otimes M_i^T M_i) w \\ &= \sum_{i=1}^r w^T (I_N \otimes M_i^T) (\mathbf{L}_i \otimes I_n) (I_N \otimes M_i) w \\ &\geq \sum_{i=1}^r \lambda_i^k ((I_N \otimes M_i) w)^T (I_N \otimes M_i) w \quad (\text{for } k = 1, 2) \\ &= \sum_{i=1}^r \lambda_i^k w^T (I_N \otimes M_i^T M_i) w \\ &= \sum_{i=1}^r \lambda_i^k w^T (I_N \otimes Q D_i) w \\ &= w^T (I_N \otimes Q \Lambda_k) w \quad [\text{by Eq. (29)}] \end{aligned}$$

Therefore

$$-w^T (I_N \otimes Q) \mathcal{L} w \leq -w^T (I_N \otimes Q \Lambda_k) w. \quad (3.36)$$

Note that the first inequality holds for  $k = 2$  by Lemma 9 and the fact that for  $x = \pi_2(u)$ , by definition,  $w^T \mathbf{1}_{nN} = 0$  and hence  $(I_N \otimes M_i) w \mathbf{1}_{nN} = 0$ . It also holds for  $k = 1$ , since  $\mathbf{L}_i$  and hence  $\mathbf{L}_i \otimes I_n$  are positive definite, and  $\lambda_i^1 = 0$ .

Now, by the Mean Value Theorem for integrals, and using (3.21), we rewrite the first term of the right hand side of (3.32) as follows:

$$\begin{aligned}
w^T (I_N \otimes Q) (\tilde{F}(u, t) - \tilde{F}(x, t)) &= \sum_{i=1}^N w^{iT} Q (F(u^i, t) - F(x^i, t)) w^i ds \\
&= \sum_{i=1}^N \int_0^1 w^{iT} Q J_F(x^i + s w^i, t) w^i ds.
\end{aligned}$$

This last equality together with (3.36) imply:

$$\begin{aligned}
&w^T (I_N \otimes Q) (\tilde{F}(u, t) - \tilde{F}(x, t)) - w^T (I_N \otimes Q) \mathcal{L}w \\
&= \sum_{i=1}^N \int_0^1 w^{iT} Q (J_F(x^i + s w^i, t) - \Lambda_k) w^i ds \\
&\leq \sum_{i=1}^N \frac{2\mu_k}{2} \int_0^1 ds w^{iT} Q w^i \\
&= \frac{2\mu_k}{2} w^T (I_N \otimes Q) w \\
&= 2\mu_k \Phi(w).
\end{aligned}$$

Therefore

$$\frac{d\Phi}{dt}(w) \leq 2\mu_k \Phi(w).$$

This last inequality implies (3.33) and (3.34) for  $k = 1$  and  $k = 2$  respectively.  $\square$

**Corollary 3** *In Theorem 3, if  $\mu_1 < 0$ , then (3.27) is contracting, meaning that solutions converge (exponentially) to each other, as  $t \rightarrow +\infty$  in the  $P$ -weighted  $L_2$  norm.*

**Corollary 4** *In Theorem 3, if  $\mu_2 < 0$ , then solutions converge (exponentially) to uniform solutions, as  $t \rightarrow +\infty$  in the  $P$ -weighted  $L_2$  norm.*

### 3.4 LMI Tests for Guaranteeing Spatial Uniformity

The next two results are modifications of Theorems 2 and 3 in [3]. They allow us to apply check the conditions in Theorems 1 and 3 through numerical tests involving linear matrix inequalities.

**Proposition 1** *If there exist constant matrices  $Z_1, \dots, Z_q$  and  $S_l, \dots, S_m$  such that for all  $x \in V$ ,  $t \in [0, \infty)$ ,*

$$J_F(x, t) \in \text{conv}\{Z_1, \dots, Z_q\} + \text{cone}\{S_l, \dots, S_m\}, \quad (3.37)$$



where

$$\text{conv}(Z_1, \dots, Z_q) = \{a_1 Z_1 + \dots + a_q Z_q \mid a_i \geq 0, \sum_i a_i = 1\},$$

and

$$\text{cone}(S_1, \dots, S_m) = \{b_1 S_1 + \dots + b_m S_m \mid b_i \geq 0\},$$

then the existence of a scalar  $\mu$  and symmetric, positive definite matrix  $Q$  satisfying

$$\begin{aligned} Q(Z_i - \Lambda_k) + (Z_i - \Lambda_k)^T Q &< \mu Q, \quad i = 1, \dots, q \\ QS_i + S_i^T Q &\leq 0, \quad i = 1, \dots, m \end{aligned} \quad (3.38)$$

implies that:

$$Q(J_F(x, t) - \Lambda_k) + (J_F(x, t) - \Lambda_k)^T Q < \mu Q \quad (3.39)$$

for all  $(x, t) \in V \times [0, \infty)$ ; or equivalently

$$\mu_k := \sup_{(x, t) \in V \times [0, \infty)} \mu_{2, P}(J_F(x, t) - \Lambda_k) < \frac{\mu}{2}, \quad (3.40)$$

where  $P^2 = Q$ .

If the image of  $V \times [0, \infty)$  under  $J_F$  is surjective onto  $\text{conv}\{Z_1, \dots, Z_q\} + \text{cone}\{S_1, \dots, S_m\}$ , then the converse is true.

*Proof* First, we rewrite the first set of conditions of (3.38) as:

$$Q\left(Z_i - \Lambda_k - \frac{\mu}{2}I\right) + \left(Z_i - \Lambda_k - \frac{\mu}{2}I\right)^T Q < 0, \quad i = 1, \dots, q \quad (3.41)$$

Defining  $D = \Lambda_k + \frac{\mu}{2}I$ , we can rewrite (3.41) as:

$$Q(Z_i - D) + (Z_i - D)^T Q < 0, \quad i = 1, \dots, q. \quad (3.42)$$

An application of [3, Theorem 2] concludes the proof. Also an application of Lemma(2) implies that (3.39) and (3.40) are equivalent.  $\square$

We define a *convex box* as:

$$\begin{aligned} \text{box}\{M_0, M_1, \dots, M_p\} &= \{M_0 + \omega_1 M_1 + \dots + \omega_p M_p \mid \omega_i \in [0, 1] \\ &\text{for each } i = 1, \dots, p\}. \end{aligned} \quad (3.43)$$

**Proposition 2** Suppose that  $J_F(x, t)$  is contained in a convex box:

$$J_F(x, t) \in \text{box}\{A_0, A_1, \dots, A_l\} \quad \forall x \in V, t \in [0, \infty), \quad (3.44)$$

where  $A_1, \dots, A_l$  are rank-one matrices that can be written as  $A_i = B_i C_i^T$ , with  $B_i, C_i \in \mathbb{R}^n$ . If there exists a scalar  $\mu$  and symmetric, positive definite matrix  $Q$  with:

$$Q = \begin{bmatrix} Q & 0 & \dots & 0 \\ 0 & p_1 & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & p_l \end{bmatrix} \quad (3.45)$$

$$Q \in \mathbb{R}^{n \times n}, \quad p_i \in \mathbb{R}, \quad i = 1, \dots, l,$$

satisfying:

$$Q \begin{bmatrix} A_0 - \Lambda_k & B \\ C^T & -I_n \end{bmatrix} + \begin{bmatrix} A_0 - \Lambda_k & B \\ C^T & -I_n \end{bmatrix}^T Q < \begin{bmatrix} \mu Q & 0 \\ 0 & 0 \end{bmatrix}, \quad (3.46)$$

with  $B = [B_1 \dots B_l]$  and  $C = [C_1 \dots C_l]$ , then the upper left (symmetric, positive definite) principal submatrix  $Q$  satisfies

$$Q(J_F(x, t) - \Lambda_k) + (J_F(x, t) - \Lambda_k)^T Q < \mu Q; \quad (3.47)$$

or equivalently

$$\mu_k := \sup_{(x,t) \in V \times [0, \infty)} \mu_{2,P}(J_F(x, t) - \Lambda_k) < \frac{\mu}{2}, \quad (3.48)$$

where  $P^2 = Q$ .

If  $l = 1$  and the image of  $V \times [0, \infty)$  under  $J$  is surjective onto  $\text{box}\{A_0, A_1\}$ , then the converse is true.

*Proof* First, we rewrite condition (3.46) as

$$Q \begin{bmatrix} A_0 - \Lambda_k - \frac{\mu}{2}I & B \\ C^T & -I_n \end{bmatrix} + \begin{bmatrix} A_0 - \Lambda_k - \frac{\mu}{2}I & B \\ C^T & -I_n \end{bmatrix}^T Q < 0. \quad (3.49)$$

Defining  $D = \Lambda_k + \frac{\mu}{2}I$ , we can rewrite (3.41) as:

$$Q \begin{bmatrix} A_0 - D & B \\ C^T & -I_n \end{bmatrix} + \begin{bmatrix} A_0 - D & B \\ C^T & -I_n \end{bmatrix}^T Q < 0. \quad (3.50)$$

An application of [3, Theorem 3] concludes the proof. Also an application of Lemma (2) implies that (3.47) and (3.48) are equivalent.  $\square$

The problem of finding the smallest  $\mu$  such that there exists a matrix  $Q$  as in Proposition 1 or a matrix  $Q$  as in Proposition 2 is quasi-convex and may be solved iteratively as a sequence of convex semidefinite programs.

**Example 4 Ring Oscillator Circuit Example**

Consider the  $n$ -stage ring oscillator whose dynamics are given by:

$$\begin{aligned} \dot{x}_1^k &= -\eta_1 x_1^k - \alpha_1 \tanh(\beta_1 x_n^k) + w_1^k \\ \dot{x}_2^k &= -\eta_2 x_2^k + \alpha_2 \tanh(\beta_2 x_1^k) + w_2^k \\ &\vdots \\ \dot{x}_n^k &= -\eta_n x_n^k + \alpha_n \tanh(\beta_n x_{n-1}^k) + w_n^k, \end{aligned} \quad (3.51)$$

with coupling between corresponding nodes of each circuit. Ring oscillators have found wide application in biological oscillators such as the repressilator in [6]. The parameters  $\eta_k = \frac{1}{R_k C_k}$ ,  $\alpha_k$ , and  $\beta_k$  correspond to the gain of each inverter. The input is given by:

$$w_i^k = d_i \sum_{j \in \mathcal{N}_{k,i}} (x_i^j - x_i^k), \quad (3.52)$$

where  $d_i = \frac{1}{R^{(i)} C_i}$  and  $\mathcal{N}_{k,i}$  denotes the nodes to which node  $i$  of circuit  $k$  is connected. We wish to determine if the solution trajectories of each set of like nodes of the coupled ring oscillator circuit given by (3.51)–(3.52) synchronize, that is:

$$x_i^j - x_i^k \rightarrow 0 \text{ exponentially as } t \rightarrow \infty \quad (3.53)$$

for any pair  $(j, k) \in \{1, \dots, N\} \times \{1, \dots, N\}$  and any index  $i \in \{1, \dots, n\}$ .

For clarity in our discussion, we take  $n = 3$  as in Fig. 3.1. We first write the Jacobian of the system (3.51), where we have omitted the subscripts indicating circuit membership:

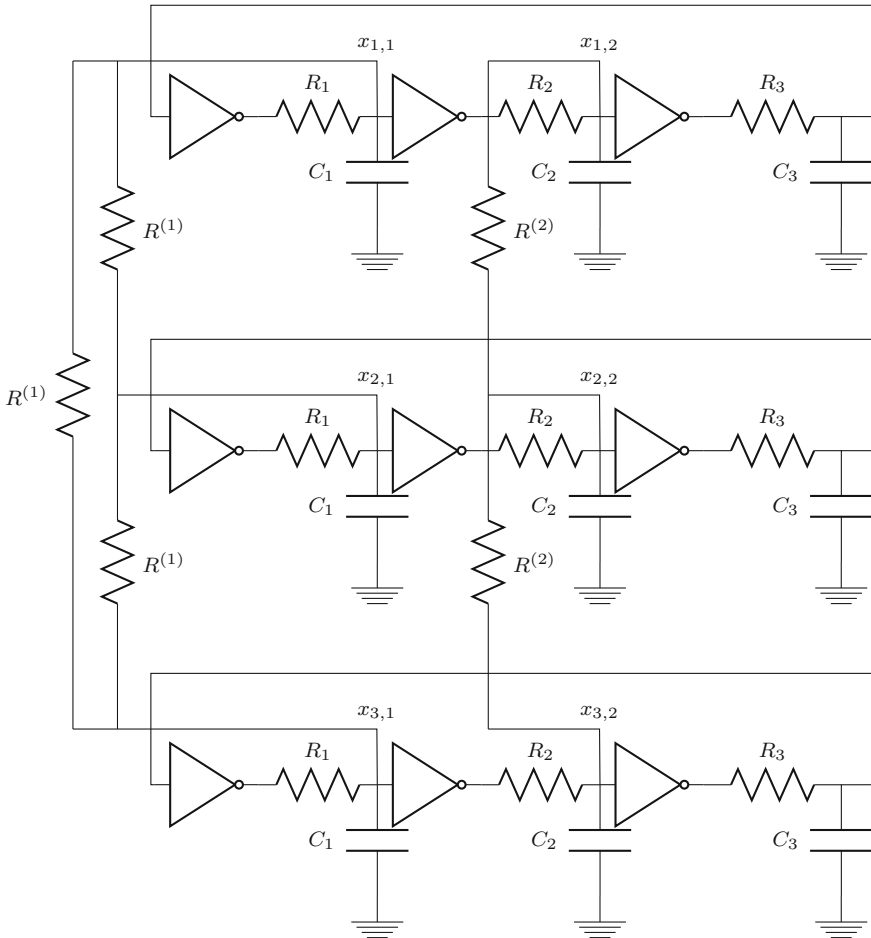
$$J(x)|_{x=\bar{x}} = \begin{bmatrix} -\eta_1 & 0 & \gamma_1(\bar{x}_1) \\ \gamma_2(\bar{x}_2) & -\eta_2 & 0 \\ 0 & \gamma_3(\bar{x}_3) & -\eta_3 \end{bmatrix}, \quad (3.54)$$

with  $\gamma_1(\bar{x}_1) = -\alpha_1 \beta_1 \operatorname{sech}^2(\beta_1 \bar{x}_3)$ ,  $\gamma_2(\bar{x}_2) = \alpha_2 \beta_2 \operatorname{sech}^2(\beta_2 \bar{x}_1)$ , and  $\gamma_3(\bar{x}_3) = \alpha_3 \beta_3 \operatorname{sech}^2(\beta_3 \bar{x}_2)$ . Define the matrices

$$\begin{aligned} A_0 &= \begin{bmatrix} -\eta_1 & 0 & 0 \\ 0 & -\eta_2 & 0 \\ 0 & 0 & -\eta_3 \end{bmatrix} & A_1 &= \begin{bmatrix} 0 & 0 & -\alpha_1 \beta_1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ A_2 &= \begin{bmatrix} 0 & 0 & 0 \\ \alpha_2 \beta_2 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & A_3 &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \alpha_3 \beta_3 & 0 \end{bmatrix}. \end{aligned} \quad (3.55)$$

Then it follows that  $J(x)$  is contained in a convex box:

$$J(x) \in \operatorname{box}\{A_0, A_1, A_2, A_3\}. \quad (3.56)$$



**Fig. 3.1** An example of a network of interconnected three-stage ring oscillator circuits as in (3.51) coupled through nodes 1 and 2

While the method of Proposition 1 involves parametrizing a convex box as a convex hull with  $2^p$  vertices, and potentially a prohibitively large linear matrix inequality computation, the problem structure can be exploited using Proposition 2 to obtain a simple analytical condition for synchronization of trajectories. In particular, the Jacobian of the ring oscillator exhibits a *cyclic* structure. The matrix  $M$  for which we seek a  $\mathcal{Q}$  satisfying (3.49), or equivalently (3.46), is given by:

$$M = \begin{bmatrix} A_0 & -\Lambda_2 & -\frac{\mu}{2}I & B \\ & C^T & & -I \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 & -\alpha_1\beta_1 \\ \alpha_2\beta_2 & 0 & 0 \\ 0 & \alpha_3\beta_3 & 0 \end{bmatrix}, \quad C = I_3. \quad (3.57)$$

Note that the matrix  $M$  exhibits a cyclic structure, and by a suitable permutation  $G$  of its rows and columns, it can be brought into a cyclic form  $\tilde{M} = G M G^T$ . Since  $\tilde{M}$  is cyclic, it is amenable to an application of the *secant criterion* [2], which implies that the condition

$$\frac{\prod_{i=1}^3 \alpha_i \beta_i}{\prod_{i=1}^3 (\eta_l + \lambda_l + \frac{\mu}{2})} < \sec^3 \left( \frac{\pi}{3} \right) \quad (3.58)$$

holds if and only if  $\tilde{M}$  satisfies

$$\tilde{Q} \tilde{M} + \tilde{M}^T \tilde{Q} < 0 \quad (3.59)$$

with negative  $\mu$ , for some diagonal  $\tilde{Q} > 0$ . Pre- and post-multiplying (3.59) by  $G^T$  and  $G$ , respectively, (3.59) is equivalent to:

$$G^T \tilde{Q} G M + M^T G^T \tilde{Q} G < 0. \quad (3.60)$$

Thus, if  $\tilde{Q}$  is diagonal and satisfies (3.59), then  $Q = G^T \tilde{Q} G$  is diagonal and satisfies (3.46). We conclude that if the secant criterion in (3.58) is satisfied, then by Proposition 2, we have:

$$\sup_{(x,t) \in V \times [0, \infty)} (J_F(x, t) - \Lambda_2) < \frac{\mu}{2}.$$

Because  $Q$  is diagonal and positive,  $Q$  is diagonal and positive. Therefore:

$$Q D_i + D_i Q > 0 \quad \text{for each } i = 1, \dots, r.$$

Therefore, since  $\mu < 0$ , by Corollary 4, we get:

$$x_i^j - x_i^k \rightarrow 0 \text{ exponentially as } t \rightarrow \infty \quad (3.61)$$

for any pair  $(j, k) \in \{1, \dots, N\} \times \{1, \dots, N\}$  and any index  $i \in \{1, 2, 3\}$ .

We note that the condition for synchrony that we have found recovers Theorem 2 in [7], which makes use of an input-output approach to synchronization [20]. We have derived the condition using Lyapunov functions in an entirely different manner from the input-output approach.

### 3.5 Conclusions

We have derived Lyapunov inequality conditions that guarantee spatial uniformity in the solutions of compartmental ODEs and reaction-diffusion PDEs even when the diffusion terms vary between species. We have used convex optimization to develop

tests using linear matrix inequalities that imply the inequality conditions, and have applied the tests to several examples of biological interest.

**Acknowledgments** Work supported in part by grants NIH 1R01GM086881 and 1R01GM100473, AFOSR FA9550-11-1-0247 and FA9550-11-1-0244, and NSF ECCS-1101876.

## References

1. Aminzare Z, Sontag ED (2012) Logarithmic Lipschitz norms and diffusion-induced instability. [arXiv:12080326v2](https://arxiv.org/abs/12080326v2)
2. Arcak M, Sontag E (2006) Diagonal stability of a class of cyclic systems and its connection with the secant criterion. *Automatica* 42(9):1531–1537
3. Arcak M (2011) Certifying spatially uniform behavior in reaction-diffusion pde and compartmental ode systems. *Automatica* 47(6):1219–1229
4. Boyd S, El Ghaoui L, Feron E, Balakrishnan V (1994) Linear matrix inequalities in system and control theory, vol 15. Society for Industrial Mathematics
5. Cantrell RS, Cosner C (2003) Spatial ecology via reaction-diffusion equations. Wiley series in mathematical and computational biology
6. Elowitz M, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403(6767):335–338
7. Ge X, Arcak M, Salama K (2010) Nonlinear analysis of ring oscillator and cross-coupled oscillator circuits. *Dyn Continuous Discrete Impulsive Syst Ser B: Appl Algorithms* 17(6):959–977
8. Gierer A, Meinhardt H (1972) A theory of biological pattern formation. *Kybernetik* 12(1):30–39
9. Gierer A (1981) Generation of biological patterns and form: some physical, mathematical, and logical aspects. *Prog Biophys Mol Biol* 37(1):1–47
10. Godsil C, Royle G, Godsil C (2001) Algebraic graph theory, vol 8. Springer, New York
11. Hale J (1997) Diffusive coupling, dissipation, and synchronization. *J Dyn Differ Equ* 9(1):1–52
12. Hartman P (1961) On stability in the large for systems of ordinary differential equations. *Can J Math* 13:480–492
13. Henrot A (2006) Extremum problems for eigenvalues of elliptic operators. Birkhauser
14. Horn RA, Johnson CR (1991) Topics in matrix analysis. Cambridge University Press, Cambridge
15. Lewis DC (1949) Metric properties of differential equations. *Am J Math* 71:294–312
16. Lohmiller W, Slotine JJE (1998) On contraction analysis for non-linear systems. *Automatica* 34:683–696
17. Lozinskii SM (1959) Error estimate for numerical integration of ordinary differential equations. *I Izv Vtssh Uchebn Zaved Mat* 5:222–222
18. Pavlov A, Pogromovsky A, van de Wou N, Nijmeijer H (2004) Convergent dynamics, a tribute to Boris Pavlovich Demidovich. *Syst Control Lett* 52:257–261
19. Russo G, di Bernardo EDSM (2010) Global entrainment of transcriptional systems to periodic inputs. *PLoS Comput Biol* 6(4)
20. Scardovi L, Arcak M, Sontag E (2010) Synchronization of interconnected systems with applications to biochemical networks: an input-output approach. *IEEE Trans Autom Control* 55(6):1367–1379
21. Schöll E (2001) Nonlinear spatio-temporal dynamics and Chaos in Semiconductors. Cambridge University Press, Cambridge
22. Smith H (1995) Monotone dynamical systems: an introduction to the theory of competitive and cooperative systems. American Mathematical Society

23. Turing AM (1952) The chemical basis of morphogenesis. *Philos Trans R Soc Lond Ser B, Biol Sci* 237(641):37–72
24. Yang XS (2003) Turing pattern formation of catalytic reaction-diffusion systems in engineering applications. *Model Simul Mater Sci Eng* 11(3):321

# Chapter 4

## Robust Tunable Transcriptional Oscillators Using Dynamic Inversion

Vishwesh V. Kulkarni, Aditya A. Paranjape and Soon-Jo Chung

**Abstract** We present a theory and associated algorithms to synthesize controllers that may be used to build robust tunable oscillations in biological networks. As an illustration, we build robust tunable oscillations in the celebrated repressilator synthesized by Elowitz and Leibler. The desired oscillations in a set of mRNA's and proteins are obtained by injecting an oscillatory input as a reference and by synthesizing a dynamic inversion based tracking controller. This approach ensures that the repressilator can exhibit oscillations irrespective of (1) the maximum number of proteins per cell and (2) the ratio of the protein lifetimes to the mRNA lifetimes. The frequency and the amplitude of at least one output (either mRNA or protein) can now be controlled arbitrarily. In addition, we characterize the  $\mathcal{L}_2$  gain stability of this 3-node network and generalize it to the case of  $N$ -node networks.

**Keywords** Transcriptional network · Elowitz-Leibler · Dynamic inversion · Adaptive control ·  $\mathcal{L}_1$  adaptive control · mRNA · Protein · Tracking controller · Stability · Zames-Falb multiplier

### 4.1 Introduction

The objective of this chapter is to illustrate how dynamic inversion control and the theory of Zames-Falb multipliers may be used to build tunable networks of synthetic biological oscillators. Synthesis of robust genetic circuits programmed to perform a particular function in vivo is a defining goal of synthetic biology. In [6], Elowitz

---

V. V. Kulkarni (✉)

University of Minnesota, Minneapolis, MN, USA

e-mail: vkulkarn@umn.edu

A. A. Paranjape · S.-J. Chung

University of Illinois at Urbana Champaign, Urbana, IL, USA

e-mail: aditya.paranjape@gmail.com

S. -J. Chung

e-mail: sjchung@illinois.edu



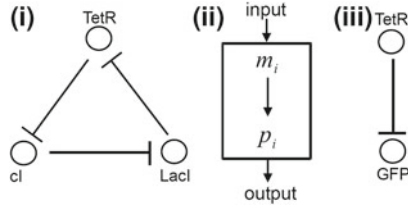
and Leibler presented one of the first synthetic biological constructs in the form of a transcriptional oscillator. This oscillator, which we shall refer to as the *EL repressilator*, is obtained by implementing a network of three non-natural transcriptional repressor systems in *Escherichia coli*. Synthesis of a green fluorescent protein is the read out for the state of this network in individual cells. Since the oscillations reported in [6], with typical periods of hours, are slower than the cell-division cycle, the state of the oscillator gets transmitted from one generation to the next. The EL repressilator was published in year 2000 and a number of interesting synthetic biological oscillators have been synthesized since then (e.g., see [1, 4, 7, 8, 13, 14, 18, 22, 23], and references therein). However, it has a compellingly simple and elegant construction, and remains a landmark in synthetic biology. Hence, we focus on the EL repressilator as the given system (i.e., the *plant* in the control theory terminology) in which tunable oscillations are to be synthesized. The controller synthesis approach is somewhat similar to the  $\mathcal{L}_1$  controller developed in [17] to induce oscillations in mitogen activated protein kinase (MAPK) cascades. This theory can be applied to synthesize robust tunable oscillations in other biological systems as well.

## 4.2 System Description

The repressilator is a cyclic negative-feedback loop comprising three repressor genes and their corresponding promoters (see Fig. 4.1). The three proteins used are LacI taken from *Escherichia coli*, TetR taken from Tn10, which is a DNA sequence with the ability to move to different positions within a single cell, and cI taken from a specific species of bacteriophage that infects *Escherichia coli*. LacI inhibits TetR transcription, TetR inhibits cI expression, and cI inhibits LacI expression, thus creating a cyclic negative feedback loop. The following first-order *ordinary differential equations* (ODEs), which assume all three repressors are identical except for their DNA, model the kinetics of the EL repressilator (see [6]):

$$\begin{aligned}\frac{dm_i}{dt} &= -m_i + \frac{\alpha_i}{1 + p_j^n} + \alpha_{i,0}, \\ \frac{dp_i}{dt} &= -\beta(p_i - m_i),\end{aligned}\tag{4.1}$$

where  $(i, j) \in \{(1, 3), (2, 1), (3, 2)\}$ , the indices 1, 2, 3 denote LacI, TetR, and cI, respectively,  $p_i$  is the concentration of the repressor-proteins,  $m_i$  is the concentration of their corresponding mRNA,  $\alpha_{i,0}$  denotes the number of protein copies per cell produced from the promoter type  $i$  during continuous growth in the presence of saturating amounts of repressor and  $\alpha_i$  is the surplus in the absence of saturating amounts of repressor,  $\beta$  is the ratio of the protein decay rate to the mRNA decay rate, and  $n$  is a Hill coefficient. Here, time is rescaled in units of the mRNA lifetime, the protein concentrations are written in units of  $K_M$ , which is the number of repressors



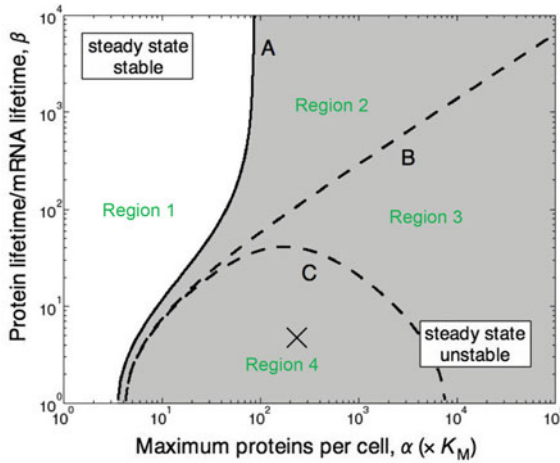
**Fig. 4.1** **i** The repressilator is a network in which three nodes are connected in cyclic inhibitory, i.e., negative, feedback loop. **ii** Each node comprises repressor a gene, its promoter, and synthesized protein. Its input stage is an mRNA  $m_i$  and its output stage is the corresponding protein  $p_i$ . This network was implemented in a plasmid by Elowitz and Leibler in [6]. **iii** The read-out reporter plasmid contains GFP fluorescence, which is inhibited by TetR. In (i) and (iii), the blunt arrows represent the inhibition interactions

necessary to half-maximally repress a promoter, and the mRNA concentrations are rescaled by their translation efficiency, which is the average number of proteins produced per mRNA molecule. In [6], it is implicitly assumed that  $\alpha_i = \alpha_j$  for all  $i, j \in \{1, 2, 3\}$  and  $\alpha_{i,0} = \alpha_0$  for all  $i \in \{1, 2, 3\}$ . Let  $\phi(x) \doteq -\frac{\alpha_i}{1+x^n}$ . Let

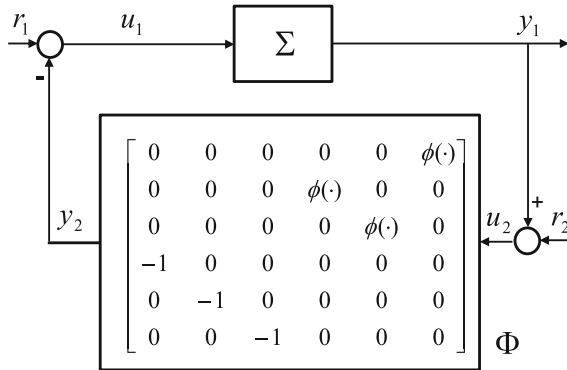
$$\Sigma \doteq \text{diag} \left( \frac{1}{s+1}, \frac{1}{s+1}, \frac{1}{s+1}, \frac{1}{s+\beta}, \frac{1}{s+\beta}, \frac{1}{s+\beta} \right).$$

Then, a block-diagram representation of the EL repressilator is as shown in Fig. 4.3; the feedback nonlinearity  $\Phi$  for this negative feedback system is specified in Fig. 4.2 while the output  $y_1$  and the exciting exogenous inputs  $r_1$  and  $r_2$  are defined as follows:  $y_1 = [m_1 \ m_2 \ m_3 \ p_1 \ p_2 \ p_3]^T$ ,  $r_1 = [\alpha_{i,0} \ \alpha_{i,0} \ \alpha_{i,0} \ 0 \ 0 \ 0]^T$ ,  $r_2 = [0 \ 0 \ 0 \ 0 \ 0 \ 0]^T$ . For mathematical convenience, we shall denote this system, i.e., the EL repressilator, as  $\mathcal{L}$ .

The stable and unstable regions in the state-space for the EL repressilator have been characterized in [6] as follows. This system of equations has a unique steady state, which becomes unstable when  $\frac{(\beta+1)^2}{\beta} < \frac{3X^2}{4+2X}$ , where  $X \doteq \frac{\alpha n p^{n-1}}{(1+p^n)^2}$  and  $p$  is a solution to the equation  $p = \frac{\alpha}{1+p^n} + \alpha_0$ . The boundary between the stable and the unstable region is as shown in Fig. 4.2. The unstable domain increases with an increase in the Hill coefficient  $n$  and is independent of  $\beta$  for sufficiently large values of  $\alpha$ . When the leakiness  $\alpha_{i,0}$  becomes comparable to  $K_M$  (which is normalized to 1 in [6]), the unstable domain shrinks (compare the curve B, for which  $\alpha_{i,0} = 0$ , to the curve C, for which  $\alpha_{i,0}/\alpha_i = 0.001$ ). Multipliers required to obtain sufficient conditions for the stability of such systems are derived in [15] and [16]. It may be remarked that by substituting the Zames-Falb multipliers used in [2] with these multipliers, it is possible to obtain a set of sufficient conditions under which this system will oscillate in response to the initial conditions alone.



**Fig. 4.2** The stability diagram of the repressilator model; the shaded region represents the unstable steady state while the unshaded region represents the stable steady state. Curve A, which consists of parameter values  $n = 2.1$  and  $\alpha_{i,0} = 0$ , marks the boundary between the two regions. The parameter values for curves B and C are  $n = 2$ ,  $\alpha_{i,0} = 0$  and  $n = 2$  and  $\alpha_{i,0}/\alpha_i = 10^{-3}$ , respectively. Also, shown are the four regions of interest: Region 1 through Region 4. (Image partially reproduced from [6])



**Fig. 4.3** A block diagram decomposition of the EL repressilator  $\mathcal{S}$

### 4.3 Stability Analysis for $\mathcal{S}$

The notation used is summarized in Table 4.1. We mostly follow the notation introduced in [5, 26]. We say that an operator  $F$  mapping a Hilbert space into itself is *positive* if the inner-product  $\langle x, Fx \rangle \geq 0 \forall x$  and *memoryless* if the output is independent of the time history of the input. We say that an operator  $F$  is *input-output stable* if it holds that every norm-bounded input  $x$  produces a norm-bounded output  $Fx$ . We say that an operator  $F$  is  $\mathcal{L}_2$ -stable if it is input-output stable with the norm

**Table 4.1** Notation

Symbol	Meaning
$(\mathbb{R}^+)$	Set of all (nonnegative) real numbers
$\mathbb{R}^n$	Set of all $n$ -dimensional real-valued vectors
$\mathbb{R}^{n \times m}$	Set of all $n \times m$ real-valued matrices
$\mathbb{Z}$	Set of all integers
$\mathcal{C}^1$	Class of continuously differentiable functions
$(\cdot)'$ or $(\cdot)^T$	Transpose of a vector or a matrix $(\cdot)$
$\text{skew}(\cdot)$	Skew [Hermitian] part of $(\cdot)$
$[\text{Herm}(\cdot)]$	
$\langle x, y \rangle$	$= \int_{-\infty}^{\infty} y^T(t)x(t)dt$
$\langle x, y \rangle_t$	$= \int_0^t y^T(t)x(t)dt$
$\ x\ $	$= \sqrt{\langle x, x \rangle}$ ( $\mathcal{L}_2$ -norm, energy of $x$ )
$\mathcal{L}_2$	Space of possibly vector valued signals $x$ for which the energy $\ x\  < \infty$ for which $\ x\ _{\mathcal{L}} < \infty \forall \mathcal{L} \in \mathbb{R}$
$\ z\ _1$	$= \int_{-\infty}^{\infty}  z(t)  dt$
$x^*$	$x^*(t) = x^T(-t)$ if $x(t) \in \mathbb{R}^n$ (we assume $x(t) = 0$ for all $t < 0$ )
$\gamma(M)$	$= \sup_x \frac{\ Mx\ }{\ x\ }$ (Gain of operator $M$ )
$\text{diag}(\alpha_i)$	Diagonal matrix with $\alpha_i$ as its elements
LTI	Linear time-invariant
SISO	Single-input-single-output
MIMO	Multi-input-multi-output

chosen to be the  $\mathcal{L}_2$ -norm. The term *multiplier* denotes a convolution operator. We say that a matrix is *stable* if the real parts of all its eigenvalues are negative valued.

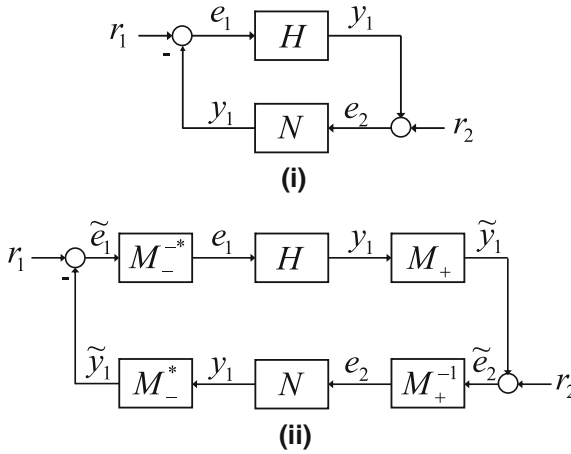
**Definition 1** A function  $f : \mathbb{R}^n \mapsto \mathbb{R}^n$  is said to be *continuously (smoothly) differentiable* if the derivative exists and is continuous (smooth).  $\square$

**Definition 2** A function  $f : \mathbb{R}^n \mapsto \mathbb{R}^m$  is said to be *Lipschitz* if there exists a constant  $L > 0$  such that, for all  $x_1, x_2 \in \mathbb{R}^n$ ,  $\|f(x_1) - f(x_2)\| \leq L\|x_1 - x_2\|$ .  $\square$

**Definition 3** A system is said to be  $\mathcal{L}_2$  stable if the energy of its output is finite for every finite energy input.  $\square$

**Definition 4** An operator  $F : \mathcal{L}_2 \longrightarrow \mathcal{L}_2$  is said to be *positive* if  $\langle x, Fx \rangle \geq 0 \forall x \in \mathcal{L}_2$ . If, additionally, there exists a constant  $\delta > 0$  such that  $\langle x, Fx \rangle \geq \delta\|x\|^2 \forall x \in \mathcal{L}_2$ , then  $F$  is said to be *strongly positive*.  $\square$

**Definition 5** Let  $A \doteq \text{diag}(a_i)$  and  $B \doteq \text{diag}(b_i)$  be matrices in  $\mathbb{R}^{n \times n}$  with  $b_i \geq a_i \forall i$ . A MIMO nonlinearity  $N : \mathcal{L}_2 \rightarrow \mathcal{L}_2$  is said to be a sector  $[A, B]$  nonlinearity if  $\langle N(x) - Ax, Bx - N(x) \rangle \geq 0 \forall x \in \mathbb{R}^n$ .  $\square$



**Fig. 4.4** **i** The feedback system  $\mathcal{E}$ :  $H$  is stable and linear time-invariant,  $N \in \mathcal{N}$ . **ii** The system  $\mathcal{E}$  after a multiplier transformation—if the Zames-Falb multipliers are used, then  $M_+$ ,  $M_+^{-1}$ ,  $M_-^*$ ,  $M_-^{*-1}$  are causal and stable with finite gain. (This figure is reproduced, in part, from [27])

**Definition 6**  $\mathcal{N}$  denotes the class of MIMO nonlinearities  $N : \mathcal{L}_2 \rightarrow \mathcal{L}_2$  for which the following equations hold.

$$\begin{aligned}
 & N(0) = 0; \\
 & \langle r - s, N(r) - N(s) \rangle \geq 0 \quad \forall r, s \in \mathcal{L}_2; \\
 & \exists \text{ a constant } c \geq 0 \text{ such that } \|N(r)\| \leq c\|r\| \quad \forall r \in \mathcal{L}_2.
 \end{aligned} \tag{4.2}$$

$$\mathcal{N}_{\text{odd}} \doteq \{N \in \mathcal{N} : N(x) = -N(-x) \quad \forall x \in \mathcal{L}_2\}. \quad \square$$

*Remark 1* The inequality (4.2) is the incremental positivity condition, which is a MIMO extension of the SISO monotonicity property. □

*Remark 2* A negative feedback interconnection of a stable LTI system and a  $\mathcal{N}$  nonlinearity is commonly referred to as a *Lure’ system* (see [21, 27]). □

A celebrated result on the  $\mathcal{L}_2$ -stability analysis of Lure’ systems is due to Zames and Falb (see [27, Theorem 2]). In [27], Zames and Falb introduced a class of non-causal multipliers to investigate the  $\mathcal{L}_2$  stability of the feedback system  $\mathcal{E}$ , shown in Fig. 4.4i, which has a stable, *linear time-invariant* (LTI) plant  $H$  in the feed-forward path and a memoryless, norm-bounded, monotone nonlinearity  $N$  in the feedback path; these multipliers are now commonly referred to as the *Zames-Falb multipliers*. The Zames-Falb multiplier approach to determining stability of a system relies on a class of possibly non-causal, linear-time-invariant multipliers that preserves the positivity of  $N \in \mathcal{N}$ . Furthermore, it is required that any multiplier  $M$  in this class be factorizable as  $M = M_- M_+$ , where  $M_-$  and  $M_+$  have the following properties:

1.  $M_-$ ,  $M_+$  are invertible; and

2.  $M_+$ ,  $M_+^{-1}$ ,  $M_-^*$ ,  $M_-^{*-1}$  are causal and have finite gain.

These properties ensure that for any such multiplier, stability of the system shown in Fig. 4.4i is equivalent to that of the system shown in Fig. 4.4ii. Stability of the system then follows if  $MH$  is strongly positive and  $N$  has finite gain (see [27, Theorem 2]).

**Theorem 1** [27, Zames-Falb Stability Theorem]

Suppose there is a mapping  $M$  (the multiplier) of  $\mathcal{L}_2$  into  $\mathcal{L}_2$  such that:

1. there are maps  $M_+$  and  $M_-$  of  $\mathcal{L}_2$  into  $\mathcal{L}_2$  with the following properties:
  - a.  $M = M_-M_+$ ;
  - b.  $M_-$  and  $M_+$  are invertible;
  - c.  $M_+$ ,  $M_+^{-1}$ ,  $M_-^*$  and  $M_-^{*-1}$  are causal and have finite gains;
2.  $MH$  and  $M^*N$  are positive;
3. either  $MH$  is strongly positive and  $H$  has a finite gain or  $M^*N$  is strongly positive and  $N$  has a finite gain.

Then  $e_1, e_2 \in \mathcal{L}_2$ . □

Thus, the key step in multiplier theoretic stability analysis of  $\mathcal{E}$  is the characterization of multipliers that preserve the positivity of the nonlinearity  $N$  of interest.

**Definition 7**  $\mathcal{M}_{odd}$  denotes the class of MIMO transfer functions (convolution operators)  $M : x \mapsto m * x$  where  $\widehat{m}(j\omega) \doteq m_0 - \widehat{z}(j\omega) \forall \omega$  and  $m_0 - \|z\|_1 > 0$ . The subclass obtained under the restriction  $z(t) \geq 0 \forall t$  is designated  $\mathcal{M}$ . The elements of  $\mathcal{M}$  and  $\mathcal{M}_{odd}$  are called the *Zames-Falb multipliers*. □

A multiplier  $M$  is said to be *positivity preserving* for a nonlinearity  $N$  if the positivity of  $N$  implies the positivity of  $MN$ . The following positivity preservation result is well known.

**Theorem 2** ([21, Theorem 2])

Suppose  $N \in \mathcal{N}$ ,  $N \in \mathcal{C}^1$  (or  $N \in \mathcal{N}_{odd}$ ,  $N \in \mathcal{C}^1$ ). Then,  $M^*N$  is positive for all  $M \in \mathcal{M}$  (or, respectively,  $M \in \mathcal{M}_{odd}$ ) if and only if  $\text{skew}(\frac{\partial N(\xi)}{\partial \xi}) = 0 \forall \xi \in \mathbb{R}^n$ . □

We shall first establish how these background results can be built upon to determine the  $\mathcal{L}_2$  stability of the EL repressilator  $\mathcal{S}$ . Thereafter, we will show how the stability analysis serves as a basis to synthesize the tunable oscillations in  $\mathcal{S}$ . Let us consider the system  $\tilde{\mathcal{S}}$  obtained from  $\mathcal{S}$  by replacing  $\phi(\cdot)$  with  $\tilde{\phi}(\cdot)$ , where  $\tilde{\phi}(x) \doteq \alpha - \frac{\alpha}{1+x^n}$  and by setting  $r_1 = [\alpha_0 \ \alpha_0 \ \alpha_0 \ -\alpha \ -\alpha \ -\alpha]^T$ . Let  $\tilde{\Phi}$  denote the MIMO nonlinearity obtained by replacing  $\phi(\cdot)$  with  $\tilde{\phi}(\cdot)$  in  $\Phi$ .

**Lemma 1** Let  $A_1, B_1 \in \mathbb{R}^{6 \times 6}$ . Let  $A_1 \doteq -I$ , where  $I$  is an identity matrix and let  $B_1 \doteq \text{diag}(\alpha)$ . Then, the nonlinearity  $\tilde{\Phi}$  is a sector  $[A_1, B_1]$  nonlinearity. □

*Proof* Observe that  $\tilde{\phi}$  is a positive valued monotone nonlinearity with gain  $\alpha$ —the gain is  $\alpha$  for infinitesimally small inputs and zero for arbitrarily large inputs. The proof follows immediately. □

The nonlinearity  $\tilde{\Phi}$  is not a positive operator. However, an equivalent feedback interconnection  $\tilde{\mathcal{S}}$  in which the feedback nonlinearity, say,  $\hat{\Phi}$  is a positive operator can be obtained from  $\tilde{\mathcal{S}}$  by applying a suitable loop-shift transformation (see [25, Chap. 5.6]).

**Lemma 2** Consider  $\tilde{\Phi}$ . The loop-shift transformation given by

$$\begin{aligned}\hat{\Phi} &\doteq (1 + (\tilde{\Phi} - A_1)((B_1 - A_1 - \delta I)^{-1}))^{-1}(\tilde{\Phi} - A_1), \\ \hat{\Sigma} &\doteq (1 + \Sigma A_1)^{-1}\Sigma + (B_1 - A_1 - \delta I)^{-1},\end{aligned}$$

where  $\delta > 0$  is arbitrarily small, transforms  $\tilde{\Phi}$  into a  $\mathcal{N}$  nonlinearity.  $\square$

*Proof* The loop-shifted system is shown in Fig. 4.5. The proof trivially follows by using block diagram reduction on the lines of the arguments presented in [25, pp. 224–225].  $\square$

A sufficient condition for the  $\mathcal{L}_2$  stability of the EL repressilator  $\mathcal{S}$  can now be stated as follows.

**Theorem 3** [Stability of the EL repressilator]

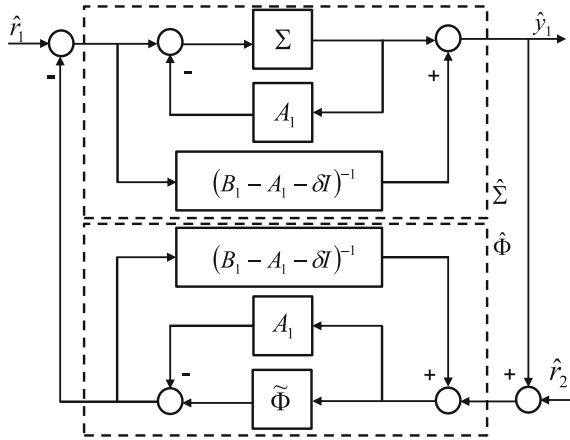
Consider the EL repressilator  $\mathcal{S}$ . Let  $A_1, B_1 \in \mathbb{R}^{6 \times 6}$ . Let  $A_1 \doteq -I$ , where  $I$  is an identity matrix and let  $B_1 \doteq \text{diag}(\alpha)$ . Let  $\hat{\Sigma} \doteq (1 + \Sigma A_1)^{-1}\Sigma + (B_1 - A_1 - \delta I)^{-1}$ . Then,  $\mathcal{S}$  is  $\mathcal{L}_2$  stable if  $\hat{\Sigma}$  is stable and if there exists an  $M \in \mathcal{M}$  such that  $M\hat{\Sigma} > 0$ .  $\square$

*Proof* Since  $\mathcal{S}$  and  $\hat{\mathcal{S}}$  are equivalent, the  $\mathcal{L}_2$  stability of  $\mathcal{S}$  is verified if the  $\mathcal{L}_2$  stability of  $\hat{\mathcal{S}}$  is verified. From Lemma 1 and Lemma 2, it follows that the feedback nonlinearity of the transformed system  $\hat{\mathcal{S}}$  is a  $\mathcal{N}$  nonlinearity. The proof then follows using Theorem 1.  $\square$

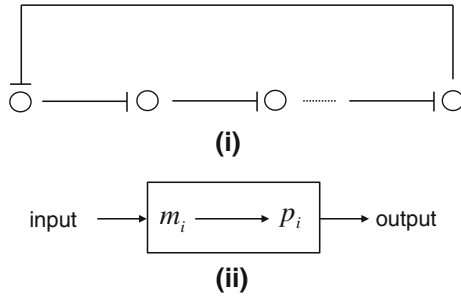
It follows that this approach scales well to cover the case of  $N\mathcal{S}$  subsystems connected in a cyclic negative feedback configuration shown in Fig. 4.6. Let us refer to this system as  $\mathcal{S}_N$ . It may be verified that Theorem 3 generalizes to such an  $N$  node network as follows.

**Theorem 4** Consider  $\mathcal{S}_N$ . Let  $A_1, B_1 \in \mathbb{R}^{2N \times 2N}$ . Let  $A_1 \doteq -I$ , where  $I$  is an identity matrix and let  $B_1 \doteq \text{diag}(\alpha)$ . Let  $\hat{\Sigma} \doteq (1 + \Sigma A_1)^{-1}\Sigma + (B_1 - A_1 - \delta I)^{-1}$ . Then,  $\mathcal{S}$  is  $\mathcal{L}_2$  stable if  $\hat{\Sigma}$  is stable and if there exists an  $M \in \mathcal{M}$  such that  $M\hat{\Sigma} > 0$ .  $\square$

We now address the problem of synthesizing tunable oscillations in the EL repressilator  $\mathcal{S}$ . We shall demonstrate how a tracking controller may be used to build robust tunable oscillations in  $\mathcal{S}$ .



**Fig. 4.5** The feedback system  $\hat{\mathcal{S}}$ . This system is obtained by loop-shifting  $\tilde{\mathcal{S}}$ . The exogenous signals and the internal signals get transformed likewise. As a result of the loop-shift, the feedback nonlinearity  $\hat{\Phi}$  is now a positive monotone nonlinearity the positivity of which is preserved by the Zames-Falb multipliers



**Fig. 4.6** **i** The network  $\mathcal{S}_N$  has  $N$  nodes that are connected in a cyclic inhibitory, i.e., negative, feedback loop. **ii** Each node comprises repressor a gene, its promoter, and synthesized protein Its input stage is an mRNA  $m_i$  and its output stage is the corresponding protein  $p_i$ . In **(i)**, the blunt arrows represent the inhibition interactions

### 4.4 Background Results for DI Controllers

Several results on synchronization of coupled oscillators have already been established (see, e.g., [2, 3, 9, 10], and references therein). The tracking controller proposed by us is based on the *dynamic inversion* (DI) theory presented in [11]. We shall demonstrate that it is equivalent to a *proportional + integral* (PI) controller with an initial condition dependent bias term. This controller ensures that the tracking error is of the same order of magnitude as the inverse of the proportional gain. It follows that a system equipped with a DI-based controller, or a PI controller in general, would respond with a periodic behaviour when an oscillatory reference signal is supplied to it.



We shall first establish a specialized version of [11, Theorem 2] for a first order system. This result is not original—this version has been described in detail in [19] and, in brief, in [20]. It is presented here for the sake of completeness and as a necessity to adequately describe the DI controller for the EL repressilator. Consider a system described by

$$\dot{x}(t) = f(x(t), z(t), u(t)), \quad \dot{z}(t) = \zeta(x(t), z(t), u(t)), \quad (4.3)$$

where  $x(0) = x_0$  and  $z(0) = z_0$  for  $(x, z, u) \in D_x \times D_z \times D_u$  and where  $D_x, D_z, D_u \subset \mathbb{R}$  are domains containing the origin. The functions  $f, \zeta: D_x \times D_z \times D_u \rightarrow \mathbb{R}$  are continuously differentiable with respect to their arguments, and furthermore, assume that  $\partial f/\partial u$  is bounded away from zero in the compact set  $\Omega_{x,z,u} \subset D_x \times D_z \times D_u$  of possible initial conditions, i.e., there exists  $b_0 > 0$  such that  $|\partial f/\partial u| > b_0$ .

Let  $e(t) = x(t) - r(t)$  be the tracking error signal. Then, the open loop error dynamics are given by

$$\begin{aligned} \dot{e}(t) &= f(e(t) + r(t), z(t), u(t)) - \dot{r}(t), \quad e(0) = e_0, \\ \dot{z}(t) &= \zeta(e(t) + r(t), z(t), u(t)), \quad z(0) = z_0. \end{aligned} \quad (4.4)$$

We construct an approximate dynamic inversion controller:

$$\varepsilon \dot{u}(t) = -\text{sign} \left( \frac{\partial f}{\partial u} \right) \mathbf{f}(t, x, z, u), \quad (4.5)$$

where

$$\mathbf{f}(t, x, z, u) \doteq f(e(t) + r(t), z(t), u(t)) - \dot{r}(t) - a_m e(t), \quad (4.6)$$

where  $a_m > 0$  gives the desired rate of convergence.

Let  $u(t) = h(t, e, z)$  be an isolated root of  $\mathbf{f}(t, e, z, u) = 0$ . The reduced system for the dynamics in (4.4) is given by

$$\begin{aligned} \dot{e}(t) &= -a_m e(t), \quad e(0) = e_0 \\ \dot{z}(t) &= \zeta(e(t) + r(t), z(t), h(t, e(t), z(t))), \quad z(0) = z_0. \end{aligned}$$

The boundary layer system is

$$\frac{dv}{d\tau} = -\text{sign} \left( \frac{\partial f}{\partial \tau} \right) \mathbf{f}(t, e, z, v + h(t, e, z)). \quad (4.7)$$

We assume that three conditions hold for all  $[t, e, z, u - h(t, e, z), \epsilon] \in [0, \infty) \times D_{e,z} \times D_v \times [0, \epsilon_0]$  for some domains  $D_{e,z}, D_v \subset \mathbb{R}$  which contain the origin:

1. The functions  $f, \zeta$  are such that their partial derivatives with respect to  $(e, z, u)$ , and the partial derivative of  $f$  with respect to  $t$  are continuous and bounded

- on any compact subset of  $D_{e,z} \times D_v$ . Further,  $h(t, e, z)$  and  $\frac{\partial f}{\partial u}(t, e, z)$  have bounded first derivatives with respect to their arguments, and  $\frac{\partial f}{\partial e}$  and  $\frac{\partial f}{\partial z}$  are Lipschitz in  $e$  and  $z$  uniformly in  $t$ .
2. The origin is an exponentially stable equilibrium of  $\dot{z}(t) = \zeta(x, z, h(t, 0, z))$ .
  3. The term  $\left| \frac{\partial f}{\partial u} \right|$ , is bounded away from zero.

**Theorem 5** ([11, Theorem 2])

Consider the boundary layer system (4.7). Suppose the above three assumptions hold. Then the origin is an exponentially stable equilibrium. Furthermore, let  $\Omega_v$  be a compact subset of  $R_v$ , where  $R_v \subset D_v$  denotes the region of attraction of the autonomous system.

$$\frac{dv}{d\tau} = -\text{sign} \left( \frac{\partial f}{\partial u} \right) \mathbf{f}(0, e_0, z_0, v + h(0, e_0, z_0)).$$

Then for each compact subset  $\Omega_{z,e} \subset D_{z,e}$  there exists a positive constant  $\epsilon_*$  and  $T > 0$  such that for all  $t \geq 0$ ,  $(e_0, z_0) \in \Omega_{e,z}$ ,  $u_0 - h(0, e_0, z_0) \in \Omega_v$ , and  $0 < \epsilon < \epsilon_*$ , the system (4.3), (4.5) has a unique solution  $x_\epsilon(t)$  on  $[0, \infty)$  and  $x_\epsilon(t) = r(t) + \mathcal{O}(\epsilon)$  holds uniformly for  $t \in [T, \infty)$ .  $\square$

*Remark 3* A DI based controller does not take advantage of any features of the system dynamics which may induce oscillatory behaviour. Therefore, a DI-based controller should be employed when the system structure does not naturally permit, or if it inhibits, an oscillatory behaviour  $\square$

*Remark 4* A DI-based controller may require high gains if small error margins are required. On such occasions, a filtered controller may have to be used. A filtered controller may potentially worsen the error margins, but can be designed to ensure stability as well as robustness. A promising approach, based on a disturbance observer, can be found in [12].  $\square$

*Remark 5* If each subsystem in  $\Sigma_1$  can be made to oscillate, the phase difference between the oscillations need not be enforced directly. Instead, the interconnection gains can be chosen to ensure a desired phase difference [24].  $\square$

## 4.5 DI Controller for Tunable Oscillations $\Sigma_1$

### 4.5.1 Problem Formulation

Consider the system of Eq. (4.1) defining the EL repressilator:

$$\begin{aligned} \frac{dm_i}{dt} &= -m_i + \frac{\alpha}{1 - p_j^n} + \alpha_{0i}, \\ \frac{dp_i}{dt} &= -\beta(p_i - m_i), \quad (i, j) = \{(1, 3), (2, 1), (3, 2)\}. \end{aligned}$$

We refer to each set of the 3 sets of equations ( $i \in \{1, 2, 3\}$ ) as a *component* of the E-L oscillator. The control inputs  $\alpha_{0i}$  can be used to ensure that  $m_i$  follows the desired reference trajectory. In particular, the frequency of the oscillations in  $m_i$  can be controlled and the phase difference between  $m_i$  and  $m_j$  can be controlled as well. This is a brute force approach to controlling the phase difference between the oscillators; in contrast, a method to choose the interconnection network between the components to achieve the desired phase difference is described in [24]. It may be noted that since the three components are structurally identical, it is, in principle, possible to send the same control signal (i.e.,  $\alpha_{0i} \equiv \alpha_0 \forall i$ ) to all components in order to ensure that the trajectories of  $m_1$ ,  $m_2$ , and  $m_3$  converge. For example, suppose  $\alpha_0$  is chosen to ensure that  $m_1$  tracks the desired reference trajectory. In this case, rigorous linear analysis readily shows that  $m_2$  and  $m_3$  converge to  $m_1$ ; MATLAB simulation results confirming this convergence are shown in Figs. 4.7 and 4.8, and we now describe the synthesis of such a controller.

### 4.5.2 Equivalence of DI and PID Control

We propose a DI based controller to induce oscillations in each oscillator. We develop a controller for each oscillator, and the same control signal is passed to the other components of the system. The control law is developed hereafter using the symbols  $x$ ,  $z$  and  $u$  for brevity of notation. Consider an *linear time-varying* (LTV) system of the form

$$\dot{x}(t) = -a(t)x(t) + \sigma(t) + g(z)u(t), \quad (4.8)$$

where  $z$  represents external dynamics which are *bounded input bounded output* (BIBO) stable with respect to  $x$ . Noise and external disturbances are captured in the term  $\sigma(t)$  which is assumed to be bounded with a bounded derivative. The control objective is to design  $u(t)$  to ensure that  $x(t)$  oscillates when both  $a(t)$  and  $g(z)$  are unknown. We assume that  $g(z) > 0 \forall z$ , i.e., the control effectiveness is positive, and that  $g(z)$  is smoothly differentiable for  $z > 0$  with a bounded derivative.

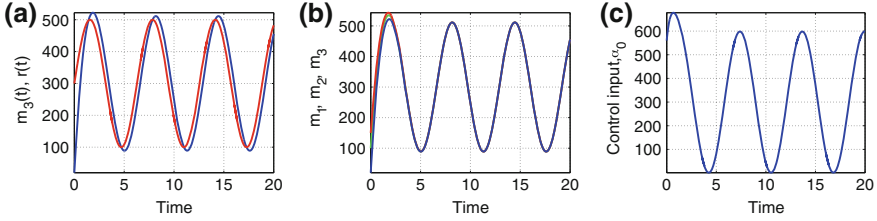
Let  $r(t)$  denote the reference signal which needs to be tracked by a given component of the EL repressilator, where  $r(t)$  can be chosen as a sine wave with an appropriate phase for each component. Then, we write the error dynamics for  $e = x - r$ :

$$\dot{e} = -a_m e + \sigma + g(z)u - a_m r - \dot{r} + (a_m - a)x, \quad (4.9)$$

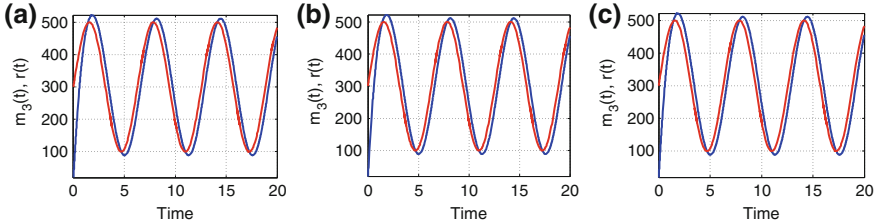
where  $a_m > 0$  ensures a desired convergence rate; in this equation, we have not stated the dependence on  $t$  explicitly. If we could ensure that  $g(z)u(t) + \sigma(t) - a_m r(t) - \dot{r}(t) + (a_m - a(t))x(t) = 0$ , then  $x(t)$  would be oscillatory. A DI-based control law given by

$$\dot{u}(t) = -k(g(z)u(t) - a_m r(t) - \dot{r}(t) + (a_m - a(t))x(t)) \quad (4.10)$$

ensures tracking with an error bounded above by  $\mathcal{O}(1/k)$ . However, note that  $a(t)$ ,  $\sigma(t)$  and  $g(z)$  are all unknown. The standard practice would be to design



**Fig. 4.7** Evolution of the states  $m_1$ ,  $m_2$ ,  $m_3$  of the three oscillators for  $\alpha = \beta = 100$  (Region 2). **a** Evolution of  $m_3$  and  $r$  with time. The output  $m_3$  tracks the reference input with a non-zero time lag and oscillates with the desired amplitude and frequency. **b** Evolution of  $m_1$ ,  $m_2$ ,  $m_3$  with time. All three outputs exhibit oscillations. Since we have only one reference input, the oscillations in these three outputs cannot be guaranteed to differ on amplitudes and frequencies. **c** Evolution of the control signal  $\alpha_0(t)$  with time



**Fig. 4.8** Evolution of the states  $m_1$ ,  $m_2$ ,  $m_3$  of the three repressilator when the parameters are chosen from the other three regions. **a** The parameters are chosen from Region 1. **b** The parameters are chosen from Region 3. **c** The parameters are chosen from Region 4. The simulation result indicate that the dynamic inversion controller is able to synthesize similar oscillations regardless of the choice of parameters

an adaptive law to estimate them [12]. Instead, using Eq. (4.9), we can rewrite the control law as

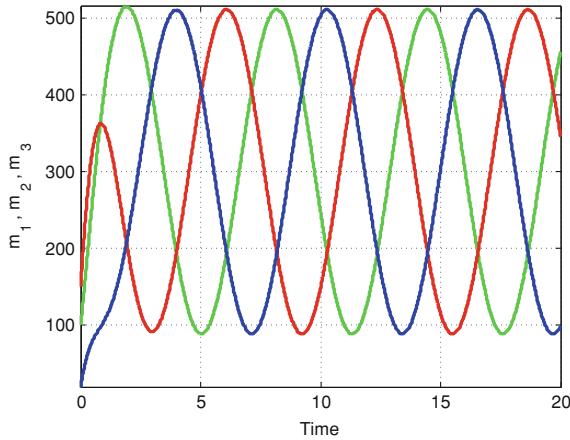
$$\dot{u}(t) = -k(\dot{e}(t) + a_m e(t)). \quad (4.11)$$

This control law still requires that we measure  $\dot{e}(t)$ . This can be done in practice by passing the signal  $e(t)$  through a lead compensator. However, as an alternative, by integrating both sides, we obtain a standard PI controller augmented by a constant which depends on the initial conditions:

$$u(t) = -k_p e(t) - k_i \int_0^t e(\tau) d\tau + u(0) + k_p e(0), \quad (4.12)$$

where  $k_p = k$ , i.e., the error bound is on the same order as the inverse of the proportional gain.

Consider the Eq. (4.12). We can set  $u(0) = 0$ . Therefore, the controller only needs to memorize  $\int_0^t e(\tau) d\tau$  and  $e(0)$ . Furthermore, the expression for  $\dot{u}(t)$  sug-



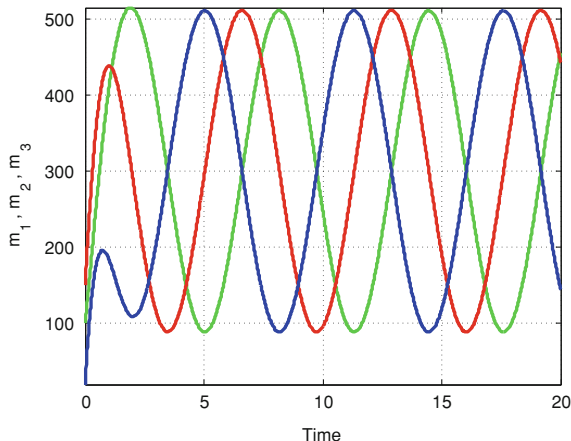
**Fig. 4.9** Evolution of the states of the three oscillators with time for  $\alpha = 200$ ,  $\beta = 3$ . Here, the commanded phase lag between the reference signal and  $m_i$  is, respectively,  $0^\circ$ ,  $120^\circ$ ,  $240^\circ$ , where  $i \in \{1, 2, 3\}$ . The plots show that the tracking controller performs well and ensures zero steady state error

gests that  $e(0)$  need not be recorded if  $\dot{e}(t)$  is available. In practice,  $\dot{e}(t)$  can be measured by passing the signal  $e(t)$  through a lead compensator which, in turn, only requires a knowledge of  $e(t)$  and  $\int_0^t e(\tau)d\tau$ . Finally, the integral action needs to be accompanied by filtering in practice to ensure that noise does not manifest itself in the integrated value. This establishes the equivalence of this DI controller and the standard *proportional integral derivative* (PID) controller. Finally, the control law designed here does not, in any way, drive the  $z$ -dynamics. Instead, the fact that they are  $\mathcal{L}_2$ -stable ensures that  $z$  does not diverge. As a result,  $g(z)$  does not diverge, it being globally Lipschitz in  $z$ .

## 4.6 Simulation Results

The simulation results for the EL repressilator  $\mathcal{S}$  when this control law is implemented are shown in Fig. 4.7 for the case of  $\alpha = \beta = 100$ —in this region,  $\mathcal{S}$  is known to be unstable (see Fig. 4.2). The control signal  $\alpha_0$  is chosen to ensure that  $m_3$  tracks the desired sinusoidal profile. The three simulation plots illustrate tunable oscillations that are generated when the parameters are chosen from the Region 2 of Fig. 4.7. The oscillations can be generated if the parameters are chosen from other regions as well (see Fig. 4.8). This illustrates the utility of our DI controller in synthesizing tunable oscillations in the EL repressilator.

If  $\alpha_{0i}$  are chosen independently of each other, it is possible to control the phase difference between the three components by sending reference signals which are offset from each other by the desired phase difference. We implemented this feedback



**Fig. 4.10** Evolution of the states of the three oscillators with time for  $\alpha = 200$ ,  $\beta = 3$ . Here, the commanded phase lag between the reference signal and  $m_i$  is, respectively,  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , where  $i \in \{1, 2, 3\}$ . The plots show that the tracking controller performs well and ensures zero steady state error

system in MATLAB. The simulation plots are presented in Figs. 4.9 and 4.10, and demonstrate that the tracking controller performs well and ensures zero steady-state error even when the desired phase differences between  $m_i$  are chosen arbitrarily.

## 4.7 Conclusion

We have presented a theory and associated algorithms to synthesize controllers that may be used to build robust tunable oscillations in the repressilator, referred to as the EL repressilator in this manuscript, synthesized in [6]. The EL repressilator is a 3-node network. We have shown how the Zames-Falb multipliers can be used to determine the  $\mathcal{L}_2$  stability of such a network. We have generalized this stability result to cover the case of  $N$ -node networks exhibiting a similar cyclic inhibitory feedback. Then, we show that the desired oscillations in a set of mRNA's and proteins can be obtained by using an oscillatory input as a reference and by synthesizing a dynamic inversion based tracking controller. This approach ensures that the repressilator can exhibit oscillations irrespective of (1) the maximum number of proteins per cell and (2) the ratio of the protein lifetimes to the mRNA lifetimes. The frequency and the amplitude of at least one output (either mRNA or protein) can now be controlled arbitrarily. The price paid for this flexibility is that we need a mechanism to set the desired reference input in the given system.

**Acknowledgments** We thank Prof. Michael Elowitz (California Institute of Technology) for clarifying our doubts on the ODE model of the EL repressilator. This research is supported, in part, by the NSF CAREER Award 0845650, NSF BIO Computing, NSF Computing and

Communications Foundationsx, and the U.S. Army Research Office Award W911NF-10-1-0296. **Competing Interests:** There are none. **Author's Contributions:** VVK derived Lemmae 1 and 2, and Theorems 3 and 4. VVK, AAP, and SJC synthesized the dynamic inversion control. AAP simulated the closed loop system using MATLAB.

## References

1. Atkinson MR, Savageau MA, Meyers J, Ninfa A (2003) Development of a genetic circuitry exhibiting toggle switch or oscillatory behavior in *escherichia coli*. *Cell* 113(5)
2. Bart-Stan G, Sepulchre R (2007) Analysis of interconnected oscillators by dissipativity theory. *IEEE Trans Autom Control* 52(2):256–270
3. Chung SJ, Slotine JJE (2010) On synchronization of coupled Hopf-Kuramoto oscillators with phase delays. *IEEE conference on decision and control*. Atlanta, GA, pp 3181–3187
4. Danino T, Mondragon-Palomino O, Tsimring L, Hasty J (2010), A synchronized quorum of genetic clocks. *Nature* 463
5. Desoer C, Vidyasagar M (1975) *Feedback systems: input-output properties*. Academic Press, New York
6. Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nat Lett* 403
7. Fung E, Wong WW, Suen JK, Bulter T, Lee S, Liao J (2005) A synthetic gene-metabolic oscillator. *Lett Nat* 435
8. Garcia-Ojalvo J, Elowitz MB, Strogatz SH (2004) Modeling a multicellular clock: repressilators coupled by quorum sensing. *Proc National Acad Sci* 101
9. Hamadeh AO, Stan GB, Goncalves JM (2010) Constructive synchronization of networked feedback systems. *IEEE conference on decision and control*. Atlanta, GA, pp 6710–6715
10. Hamadeh AO, Stan GB, Sepulchre R (2012) Global state synchronization in networks of cyclic feedback systems. *IEEE Trans Autom Control* 57(2)
11. Hovakimyan N, Lavretsky E, Sasane A (2007) Dynamic inversion for nonaffine-in-control systems via time-scale separation. Part i. *J Dyn Control Syst* 13(4)
12. Kharisov E, Kim K, Wang X, Hovakimyan N (2011) Limiting behavior of  $\mathcal{L}_1$  adaptive controllers. In: *Proceedings of AIAA guidance, navigation and control conference*. Portland, OR
13. Kim J, Shin D, Jung S, Heslop-Harrison P, Cho K (2010) A design principle underlying the synchronization of oscillations in cellular systems. *J Cell Sci* 123(4)
14. Kim J, Winfree E (2011) Synthetic in vitro transcriptional oscillators. *Mol Syst Biol* 7(465)
15. Kulkarni VV, Pao LY, Safonov MG (2011) On stability analysis of systems featuring a multiplicative combination of nonlinear and linear time-invariant feedback. *Int J Robust Nonlinear Control* 21(18)
16. Kulkarni VV, Pao LY, Safonov MG (2011) Positivity preservation properties of the Rantzer multipliers. *IEEE Trans Autom Control* 56(1)
17. Kulkarni VV, Paranjape AA, Ghusinga KR, Hovakimyan N (2010) Synthesis of robust tunable oscillators using mitogen activated protein kinase cascades. *Syst Synth Biol* 4
18. Montagne K, Plasson R, Sakai Y, Rondelez Y (2011) Programming an in vitro DNA oscillator using a molecular networking strategy. *Mol Syst Biol* 7(466)
19. Paranjape AA (2011) Dynamics and control of robotic aircraft with articulated wings. Ph.D. thesis, University of Illinois at Urbana-Champaign, Champaign, IL
20. Paranjape AA, Kim J, Chung SJ (2012) Closed-loop perching of aerial robots with articulated flapping wings. *IEEE Trans Robot* (Submitted)
21. Safonov M, Kulkarni V (2000) Zames-Falb multipliers for MIMO nonlinearities. *Int J Robust Nonlinear Control* 10(11/12):1025–1038
22. Strelkova N, Barahona M (2010) Switchable genetic oscillator operating in quasi-stable mode. *J Roy Soc Interface* 7(48)

23. Tigges M, Marquez-Lago T, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. *Nature* 457(4)
24. Varigonda S, Georgiou T (2001) Dynamics of relay-relaxation oscillators. *IEEE Trans Autom Control* 46(4)
25. Vidyasagar M (1993) *Nonlinear systems analysis*, 2nd edn. Prentice-Hall, Englewood Cliffs
26. Willems J (1971) *The analysis of feedback systems*. The MIT Press, Cambridge
27. Zames G, Falb P (1968) Stability conditions for systems with monotone and slope-restricted nonlinearities. *SIAM J Control Optim* 6:89–108



# Chapter 5

## Towards the Modular Decomposition of the Metabolic Network

Anne Goelzer and Vincent Fromion

**Abstract** Modular systems emerged in biology through natural selection and evolution, even at the scale of the cell with the cellular processes performing elementary and specialized tasks. However, the existence of modules is questionable when the regulatory networks of the cell are superimposed, in particular for the metabolic network. In this chapter, a theoretical framework that allows the breakdown of the steady-state metabolic network of bacteria into elementary modules is introduced. The modular decomposition confers good systemic and control properties, such as the decoupling of the steady-state regime with respect to the co-metabolites or co-factors, to the entire system. The biological configurations and their impact on the module properties are discussed in detail. In particular, the presence of irreversible enzymes was found to be critical in the module definition. Moreover, the proposed framework can be used to qualitatively predict the dynamics of the module components and to analyse biological datasets.

**Keywords** Modular · Metabolic network · Steady-state · Metabolite · RNA · DNA · Enzyme · *Bacillus subtilis* · Genetic · Regulatory network · Pathway · Genetic control · End product control structure (EPCS)

### 5.1 Introduction

Modularity emerged at all scale in living organisms, from organs in mammals to cellular processes in bacteria such as DNA replication. These sub-systems, empirically identified through their functions, perform elementary specialized tasks, that

---

A. Goelzer · V. Fromion (✉)  
INRA, UR1077 Unité Mathématique Informatique et Génome, 78350 Jouy en Josas, France  
e-mail: vincent.fromion@jouy.inra.fr

A. Goelzer  
e-mail: anne.goelzer@jouy.inra.fr

are coordinated to achieve the growth and the survival of the organism. Despite the existence of these specialized sub-systems, the existence of modules is questionable when the regulatory networks of the cell are superimposed, and in particular for the metabolic network. The metabolic network is a central cellular process whose main function is to produce energy and the main building blocks for biomass synthesis like amino acids or nucleotides. It is composed of a large set of highly connected chemical reactions (more than 2,000 reactions for the bacterium *Escherichia coli* [11]) catalysed by enzymes. The questions that we addressed in this chapter is: can we identify modules and, more generally any structure in the metabolic network when the metabolic regulatory network is considered? Can we establish intrinsic and structural properties associated to this organisation? These questions are ambitious and require, as a preliminary step, to have the regulatory network of the metabolic pathways of an organism, enough complete and exhaustive to tackle these questions. To this purpose, we focused on the metabolic network of the “simplest” organism, the bacterium. However, since the metabolic pathways are highly conserved in higher organisms, the results obtained in this chapter are also interesting.

In previous works [17], we inferred the genetic and metabolic regulatory network for the model bacterium *Bacillus subtilis* using information in the literature and databases. From the analysis of this network, we pointed out, in agreement with the results of [21], the key role of metabolites in the genetic control of metabolic networks. Moreover, we identified (a) two main control structures of metabolic pathways and (b) the standard biological configurations that are found in the metabolic network. Most existing studies focus on the behaviour of metabolic pathways (or signalling pathways) and consider a specific metabolic configuration [3, 28, 34, 36]. Because of the strong non-linearity of the dynamical model that is used to describe the system, these authors mainly focused on identifying the stability conditions for a simplified model. Moreover, their results are rarely discussed from a biological point of view. Some work has dealt with more realistic metabolic configurations [1, 2], but these models do not integrate genetic regulation.

In contrast to these studies, our approach analyses the existence and uniqueness of a structural steady-state regime for any metabolic pathway, regardless of its configuration and its genetic and enzymatic regulatory mechanisms. We identified two types of well-defined elementary modules that have specific mathematical properties. This module definition can then be used to study the decomposition of a complete metabolic network into modules.

This chapter is organised as follows. Section 5.2 briefly introduces the main results of our work [17] and details the identification of two control structures in the metabolic network, which are considered elementary modules. Sections 5.3 and 5.4 discuss the existence and uniqueness of a steady state in the two elementary modules and in a large diversity of biological configurations. Section 5.5 examines the connection and the coordination between modules. Section 5.6 focuses on the decomposition of the metabolic network of *B. subtilis* into modules.

## 5.2 Two Main Control Structures in Metabolic Pathways

The analysis of the *B. subtilis* metabolic network (see Fig. 5.1 (top) and [17]) led to the identification of two distinct control structures in metabolic pathways. In the first control structure, which we named end-product control structure (EPCS), the last metabolite of the metabolic pathway is the key factor because it inhibits the activity of the first enzyme and its synthesis through a genetic regulator. The second structure, which is called initial-product control structure (IPCS), involves the first metabolite of the pathway. Increasing concentrations of the first metabolite induces the synthesis of the enzymes in the pathway through a genetic regulator. Based on previous results [17], we defined two levels of control in metabolic pathways: local regulation and global regulation [see Fig. 5.1 (bottom)]. The local regulation of a metabolic pathway corresponds to any type of genetic regulation (transcriptional, translational, and post-translational) that involves the concentration of an intermediate metabolite in the controlled pathway. The global regulation of a metabolic pathway is defined as all non-local regulations. For practical purposes, the local regulation ensures the induction or repression of enzymes of the pathway according to the concentration of an intermediate metabolite of the pathway. The global regulation, however, can change or bypass the local regulation.

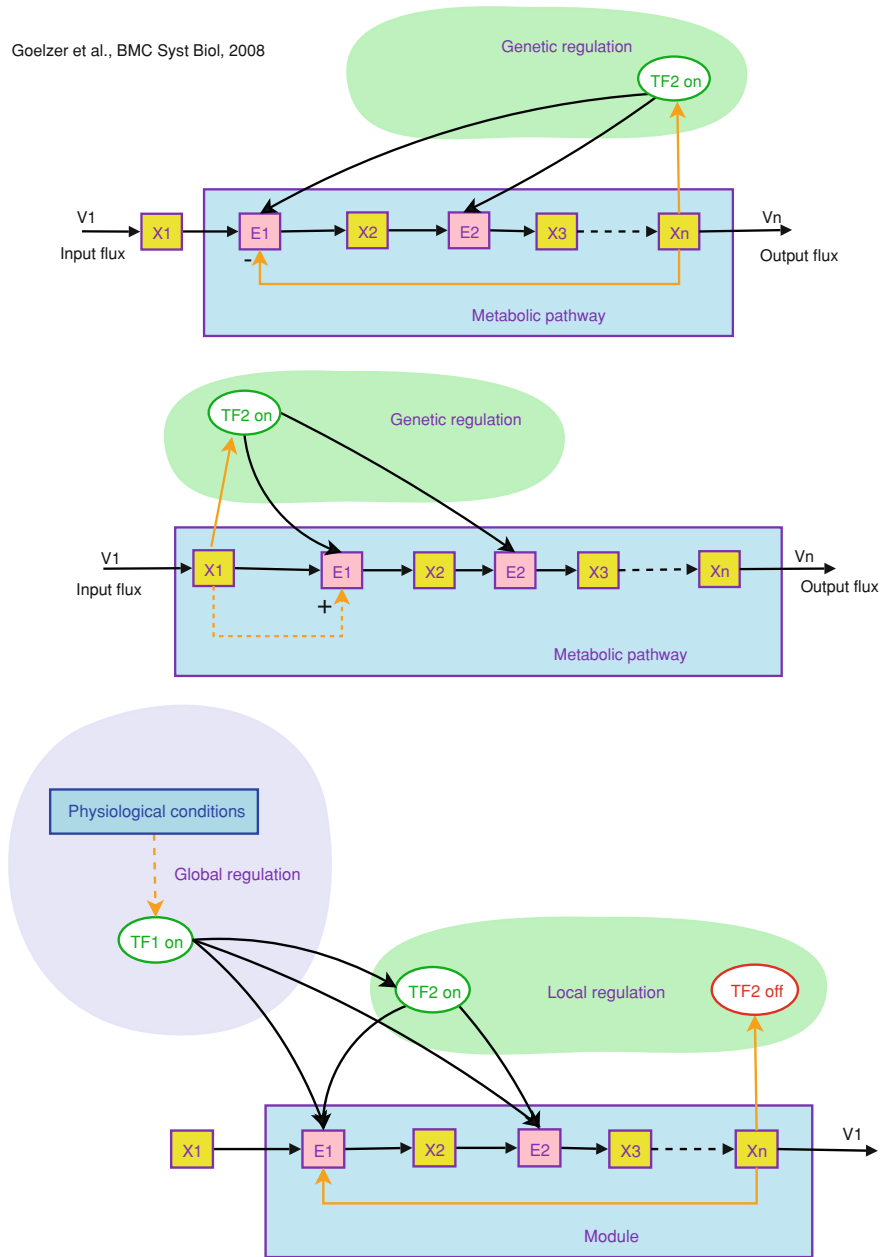
The choice of these structures as elementary sub-systems, even if it seems simple at first, is based on their intrinsic mathematical properties, which will be presented in the next sections. From an input/output perspective, these control structures correspond to sub-systems, or **modules**. In addition, these allow the breakdown of the metabolic network into sub-systems, which usually correspond to the empirical organisations of the metabolic network that are defined by biologists.

## 5.3 The End-Product Control Structure

In this section, the theoretical properties that are related to the end-product control structure are analysed. In addition, the consequences of these properties will be assessed from a biological point of view. Compared to previous studies, this work systematically studies the impact of different biological configurations of metabolic pathways, which are deduced from the work by [17]. These configurations include changes in the reversibility/irreversibility of metabolic pathways, the presence of cofactors, and isoenzymes and the organisation of the genes in an operon.

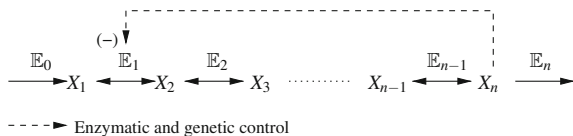
The EPCS system is shown in Fig. 5.2. As shown, the system is a linear metabolic pathway that is composed of  $n$  metabolites ( $X_1, \dots, X_n$ ) and  $n - 1$  enzymes ( $\mathbb{E}_1, \dots, \mathbb{E}_{n-1}$ ) and is controlled by the concentration of the end-product, which represses the synthesis of the first enzyme  $\mathbb{E}_1$  (genetic level) and inhibits the activity of  $\mathbb{E}_1$  (metabolic level).

Goelzer et al., BMC Syst Biol, 2008



**Fig. 5.1** Top two control structures in the metabolic network: one controlled by the last metabolite (end-product), one controlled by the first metabolite (initial-product). Enzymes (resp. metabolites) are in pink (resp. yellow). The transcription factor (TF) is the ellipsoid, and the orange arrows refer to the regulation by metabolites on the enzyme activity and on the TF activity. Bottom two control levels of a metabolic pathway

**Fig. 5.2** A metabolic pathway controlled by the end-product



**Model of metabolic level:** Following the standard representation of enzymatic reaction of [38], the dynamics of the metabolite concentrations can be described by the following set of ordinary differential equations:

$$\begin{cases} \dot{x}_1(t) = v_0(t) - E_1(t) f_1(x_1(t), x_2(t), x_n(t)) \\ \dot{x}_2(t) = E_1(t) f_1(x_1(t), x_2(t), x_n(t)) - E_2(t) f_2(x_2(t), x_3(t)) \\ \vdots \\ \dot{x}_n(t) = E_{n-1}(t) f_{n-1}(x_{n-1}(t), x_n(t)) - v_n(t) \end{cases} \quad (5.1)$$

where  $v_n(t) \triangleq E_n(t) f_n(x_n(t))$  and the characteristics of the enzyme activity  $f_i$  are such that:

(a) **Reversible enzymes:**

- for intermediate enzymes,  $\mathbb{E}_i$  for  $i \in \{2, \dots, n-1\}$ :  $f_i$  is continuous, increasing in  $x_i$  and decreasing in  $x_{i+1}$  such that  $f_i(0, 0) = 0$ ,  $f_i(x_i, 0) > 0$  for all  $x_i > 0$ , and  $f_i(0, x_{i+1}) < 0$  for all  $x_{i+1} > 0$ . Moreover, there exist  $M_i > 0$ ,  $M'_i \geq 0$ <sup>1</sup> such that  $f_i(x_i, x_{i+1}) \in (-M'_i, M_i)$  for all  $x_i > 0$  and  $x_{i+1} \geq 0$ . We assume that for  $x_i > 0$ , there always exists  $x_{i+1} > 0$  such that  $f_i(x_i, x_{i+1}) = 0$ .
- for the first enzyme,  $\mathbb{E}_1$ :  $f_1(0, 0, x_n) = 0$  for all  $x_n \geq 0$ ,  $f_1(x_1, 0, x_n) > 0$  for all  $x_1 > 0$  and  $x_n \geq 0$ , and  $f_1(0, x_2, x_n) < 0$  for all  $x_2 > 0$  and  $x_n \geq 0$ . Moreover, there exist  $M_1 > 0$ ,  $M'_1 \geq 0$  such that  $f_1(x_1, x_2, x_n) \in (-M'_1, M_1)$ , for all  $x_1 > 0$ ,  $x_2 \geq 0$  and  $x_n \geq 0$ . Moreover, we also assume that for all  $x_1 > 0$  and  $x_n \geq 0$ , there exists  $x_2 > 0$  such that  $f_1(x_1, x_2, x_n) = 0$ . In addition,  $f_1(x_1, x_2, x_n)$  is continuous and increasing (resp. decreasing) in  $x_1$  (resp.  $x_2$ ). Moreover, if  $f_1(x_1, x_2, 0) < 0$  for all  $x_1 > 0$  and  $x_2 > 0$ , then  $f_1$  is increasing in  $x_n$ ; similarly, if  $f_1(x_1, x_2, 0) > 0$  for all  $x_1 > 0$  and  $x_2 > 0$ , then  $f_1$  is decreasing in  $x_n$ . Moreover, for all  $x_1 > 0$  and  $x_2 > 0$ ,  $\lim_{x_n \rightarrow +\infty} f_1(x_1, x_2, x_n) = 0$ .
- for the last enzyme,  $\mathbb{E}_n$ :  $\mathbb{E}_n$  represents the set of chemical reactions that consume  $x_n$  and summarises the link between the flux that is produced by the metabolic pathway and the final concentration. The characteristics of  $f_n$  mainly depend on other modules. In addition,  $f_n$  is continuous and increasing in  $x_n$  such that  $f_n(0) = 0$ .

<sup>1</sup> All constants introduced in this chapter are assumed to be finite.

(b) **Irreversible enzymes:**

- for the intermediate enzymes,  $\mathbb{E}_i$  for  $i \in \{2, \dots, n-1\}$ :  $f_i$  is continuous and increasing in  $x_i$  such that  $f_i(0) = 0$ . Moreover, there exists  $M_i > 0$  such that  $f_i(x_i) \in (0, M_i)$  for all  $x_i > 0$  and  $\lim_{x_i \rightarrow +\infty} f_i(x_i) = M_i$ .
- for the first enzyme,  $\mathbb{E}_1$ :  $f_1(0, x_n) = 0$  for all  $x_n \geq 0$ . There exists  $M_1 > 0$  such that  $f_1(x_1, x_n) \in (0, M_1)$  for all  $x_1 > 0$  and  $x_n \geq 0$ . Moreover,  $f_1$  is continuous, increasing in  $x_1$  and decreasing in  $x_n$  such that for all  $x_1 > 0$ ,  $\lim_{x_n \rightarrow +\infty} f_1(x_1, x_n) = 0$ .

**Model of the control at the genetic level:** Enzyme synthesis occurs in two steps: the gene is first transcribed by the RNA polymerase to produce the RNA messenger (mRNA), which is then translated by the ribosomes to produce the protein. By noting  $m$ ,  $Y_L$  and  $R_L$  as the concentrations of mRNAs, free RNA polymerases and free ribosomes, respectively, a simplified dynamic model of the synthesis of an enzyme  $\mathbb{E}$  can be written:

$$\begin{cases} \dot{m}(t) = k_m Y_L(t - \tau_m) - k_d m(t) \\ \dot{E}(t) = k_e m(t) R_L(t - \tau_e) - \mu E(t) \end{cases} \quad (5.2)$$

where

- $k_m$ ,  $k_d$  and  $k_e$  are the affinity of the promoter for the RNA polymerase, the degradation of mRNA and the affinity of the ribosome for the mRNA, respectively;
- $\mu$  is the growth rate of the bacterium in exponential growth ( $\mu$  can then be calculated such that  $\dot{N}(t) = \mu N(t)$ , where  $N(t)$  is the concentration of the bacterial population);
- $\tau_m$  is the transcriptional delay, which corresponds to the time required for mRNA availability for ribosomes; and
- $\tau_e$  is the translational delay, which corresponds to the time required for translation of the mRNA.

Moreover, if the synthesis of the mRNA or the enzyme is inhibited by a factor, such as a metabolite, the previous equations also depend on the factor. For example, if the synthesis of the mRNA is inhibited by the metabolite  $X$ , which has a concentration of  $x$ , then the first equation is now

$$\dot{m}(t) = k_m f_I(x(t)) Y_L(t - \tau_m) - k_d m(t)$$

where  $f_I(x(t))$  is continuous, positive and decreasing in  $x$ .

If the concentration of the metabolite  $x$  has a constant steady state regime  $\bar{x}$ , then

$$\bar{m} = \frac{k_m}{k_d} f_I(\bar{x}) \bar{Y}_L, \quad \bar{E} = \frac{k_e k_m}{\mu k_d} f_I(\bar{x}) \bar{Y}_L \bar{R}_L.$$







$$\bar{E}_i = \alpha_i \frac{g(\bar{x}_n)}{\mu} \quad \text{and} \quad \bar{x}_i = f_i^{-1} \left( \frac{\mu E_n f_n(\bar{x}_n)}{\alpha_i g(\bar{x}_n)} \right) \quad (5.9)$$

if and only if for  $i \in \{2, \dots, n-1\}$

$$\alpha_i M_i > \frac{\mu E_n f_n(\bar{x}_n)}{g(\bar{x}_n)}. \quad (5.10)$$

The proof of this proposition is a particular case of the proof of Proposition 3, which is shown in page 10. Proposition 1 indicates that the system (5.7) has a unique steady-state regime if and only if all of the enzymes that belong to the metabolic pathway do not saturate (the condition (5.10) holds true). Moreover,  $x_n$  and thus implicitly  $f_1$  and  $g$  have key roles in the definition of the steady state. The monotonicity of  $f_1$  and  $g$  with respect to  $x_n$  allows to deduce the unicity of  $x_n$ . Surprisingly, the characteristics denoted by  $f_i$  and the concentrations of the intermediate enzymes have no impact on the definition of the steady state  $\bar{E}_1$ ,  $\bar{x}_n$  and the output flux  $E_n f_n(\bar{x}_n)$  if none of the intermediate enzymes saturate. Consequently, the sensitivity of the steady-state regime to a constant perturbation in the concentration of enzyme  $E_n$  (or to a flux demand  $v_n$ ) only depends on the genetic characteristics  $g$  and the characteristics  $f_1$  of the first enzyme. The prediction of the steady-state behaviour of the metabolic pathway can therefore be dramatically simplified, even if it is composed of a large number of intermediate reactions.

*Remark 3* The condition (5.10) can be written as

$$\frac{\mu E_n f_n(\bar{x}_n)}{g(\bar{x}_n)} < \alpha_i M_i \iff f_1(\bar{x}_1, \bar{x}_n) < \alpha_i M_i.$$

Therefore, if  $M_1 < \alpha_i M_i$  for all  $i \in \{2, \dots, n-1\}$ , then condition (5.10) is satisfied.

### 5.3.1.2 Behaviour of the Components of the Metabolic Pathway

The variation of the flux demand with respect to the variation of the concentration of  $\mathbb{E}_n$  will now be discussed. Based on the definition of the steady-state regime,

$$\frac{f_1(\bar{x}_1, \bar{x}_n)g(\bar{x}_n)}{\mu f_n(\bar{x}_n)} \triangleq E_n.$$

Therefore,  $\bar{x}_n$  is decreasing when  $E_n$  is increasing. In addition, the final flux demand  $\bar{v}_n \triangleq E_n f_n(\bar{x}_n)$  is by definition equal to  $f_1(\bar{x}_1, \bar{x}_n)g(\bar{x}_n) = \bar{v}_n$ . Because the left side of equation is a decreasing function of  $\bar{x}_n$ , then, when  $E_n$  is increasing,  $\bar{v}_n$  is also increasing (as long as none of the enzymes saturate). Consequently, the metabolic pathway has a maximal flux capability, which is given by the following corollary.

**Corollary 1** *Let the assumptions of Proposition 1 be satisfied. Then the flux demand has the following upper bound at steady state*

$$\frac{g(0)}{\mu} f_1(\bar{x}_1, 0). \quad (5.11)$$

The outer flux is then bounded and the superior value only depends on the characteristics  $f_1$  and  $g$  of the first enzyme; this is only true if none of the intermediate enzymes saturate.

The impact of variations in (a) the flux demand and (b) the concentration  $x_1$  on the intermediate metabolite concentrations will now be discussed.

**Corollary 2** *Let the assumptions of Proposition 1 be satisfied. Then, (a) for all  $i \in \{2, \dots, n-1\}$ ,  $\bar{x}_i = \bar{x}_i(E_n)$  is increasing in  $E_n$  and  $\bar{x}_n = \bar{x}_n(E_n)$  is decreasing in  $E_n$  and, (b) for all  $i \in \{2, \dots, n\}$ ,  $\bar{x}_i(\bar{x}_1)$  and  $\bar{v}_n(\bar{x}_1)$  are increasing in  $\bar{x}_1$ .*

The intermediate metabolite concentrations are increasing functions of the flux demand and of  $x_1$ , whereas the end-product is a decreasing (resp. increasing) function of the flux demand (resp.  $x_1$ ).

*Remark 4*  $\bar{x}_n$  can be written as a function of  $\bar{x}_1$ :  $\bar{x}_n \triangleq H(\bar{x}_1)$ . Therefore, at steady state, the input and output flux and the concentration of the first metabolite  $\bar{x}_1$  are linked by the monotonously increasing relationship  $v_0 = E_n f_n(H(\bar{x}_1))$ . We then obtain an input/output description that corresponds to a fictitious enzyme, which links  $v_0$  to  $\bar{x}_1$  and integrates all of the module properties through the functions  $H$  and  $f_n$ .

*Remark 5* A metabolic flux corresponds to a material flow through an enzyme such that  $v = E f_E(x)$ . A metabolic flux is thus an intensive quantity, whereas the metabolite concentration is an extensive quantity. This fact explains why, in most mechanisms of gene regulation, only the concentration of a metabolite is used (and not the flux). As in Ohm's law ( $U = RI$ ), in which the current  $I$  is measured through the measurement of the voltage  $U$  for the resistance  $R$ , the cell senses the flux  $v$  through the measurement of the concentration  $x$  and a specific mechanism, such as an enzyme or a genetic regulator.

### 5.3.1.3 Consequences of Enzyme Saturation

Several factors can result in enzyme saturation; these include an inadequate concentration of the enzyme or its limitation by a cofactor. The effect of enzyme saturation will now be discussed.

**Corollary 3** *Let the assumptions of Proposition 1 be satisfied and let us define*

$$\psi_{i^*,sat} = \min_{i \in \{2, \dots, n-1\}} \alpha_i M_i$$

and  $i^*$ , the value of  $i$  for which the minimum is reached ( $i^*$  can also correspond to a set of possible values). If  $\psi_{i^*,sat}$  is such that  $\psi_{i^*,sat} < f_1(\bar{x}_1, 0)$ , then there exists  $E_n^*$  and  $\bar{x}_n^*$  such that

$$\frac{\mu E_n^* f_n(\bar{x}_n^*)}{g(\bar{x}_n^*)} = \psi_{i^*,sat}$$

and

$$\lim_{E_n \rightarrow E_n^*} \bar{x}_n^* = +\infty.$$

In addition, for  $E_n \geq E_n^*$ , the regime of the metabolic pathway is saturated.

The output flux is fixed through the saturation of the enzyme  $\psi_{i^*,sat}$  and by the characteristics  $g$  of  $\mathbb{E}_1$ . Moreover, the concentration of the metabolite  $\bar{x}_n^*$ , which is the substrate of enzyme  $i^*$ , goes theoretically to infinity when  $E_n$  goes to  $E_n^*$ . Obviously, thermodynamical laws prevent the metabolite concentration to go to infinity. Very high concentrations of metabolites lead to reverse the direction of the chemical reaction, i.e. the irreversible enzyme becomes reversible (see Sect. 5.3.1.6).

### 5.3.1.4 Biological Interpretation

The biosynthesis pathways of amino acids are generally regulated by the end product. The enzyme  $\mathbb{E}_n$  and the output flux  $v_n$  correspond to the tRNA synthase and the flux of charged-tRNA that is consumed by the ribosomes for the production of proteins at steady-state, respectively. Thus, an increase in the ribosomal demand usually results in an increase in the concentration of tRNA synthase ( $E_n$ ) due to a genetic regulation that induces a decrease in the concentration of the amino acid  $x_n$ . A decrease in  $x_n$  leads to the readjustment of the entire pathway (enzyme and metabolites) to provide the requested flux (assuming that the intermediate enzymes do not saturate). In other words, for fixed  $\bar{x}_1$ , the concentration of the amino acid  $x_n$  must decrease to increase the capacity of the synthesis pathway and thus satisfy the flux demand within the limit defined by the characteristics of the first enzyme (Corollary 1).

### 5.3.1.5 The Genes are Independent

In the following analysis, the genes belonging to the metabolic pathway are not in the same operon. We assume that a steady state for the intermediate enzyme exists and is given by  $(E_i)_{i \in \{2, \dots, n\}} > 0$ .

**Proposition 2** For all  $\mu > 0$ ,  $\bar{x}_1 > 0$  and  $E_i > 0$  for  $i \in \{2, \dots, n\}$ , there exists an unique steady-state regime  $\bar{E}_1$  and  $(\bar{x}_2, \dots, \bar{x}_n)$  for (5.7) such that

$$\begin{cases} \bar{E}_1 = \frac{g(\bar{x}_n)}{\mu} \\ f_1(\bar{x}_1, \bar{x}_n)g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n) \\ v_0 = E_n f_n(\bar{x}_n) \end{cases} \quad (5.12)$$

and for all  $i = \{2, \dots, n-1\}$ ,  $\bar{x}_i = f_i^{-1}\left(\frac{E_n f_n(\bar{x}_n)}{E_i}\right)$  if and only if  $E_n f_n(\bar{x}_n) < E_i M_i$ .

Compared to Proposition 1, only the condition of saturation changes. The link between the flux demand and the concentrations of the first and last metabolite that are obtained in Proposition 1 is unchanged as long as none of the intermediate enzymes saturate. All of the previous results of Sect. 5.3.1.1 can be easily extended.

### 5.3.1.6 All Enzymes are Reversible

We now assume that all of the enzymes in the metabolic pathway (including the first enzyme) are reversible. This configuration dramatically changes the properties obtained in Proposition 1. In contrast, the results in Proposition 1 can be partially recovered through the presence of a single irreversible enzyme.

**Proposition 3** *If the genes coding for  $(\mathbb{E}_1, \dots, \mathbb{E}_{n-1})$  belong to the same operon (see Eq. (5.6)) and if the enzymes  $\mathbb{E}_i$  for all  $i \in \{1, \dots, n-1\}$  are reversible, then, for all  $\mu > 0$ ,  $E_n > 0$  et  $\bar{x}_1 > 0$ , there exists a unique steady-state regime for the system (5.5),  $(\bar{E}_1, \dots, \bar{E}_{n-1})$  and  $(\bar{x}_2, \dots, \bar{x}_n)$  such that*

$$\begin{cases} \bar{x}_n = H_n(H_2(\dots(H_{n-1}(\bar{x}_n, \bar{x}_1), \bar{x}_1) \dots, \bar{x}_1), \bar{x}_1) \\ \bar{x}_i = H_i(H_{i+1}(\dots(H_{n-1}(\bar{x}_n, \bar{x}_1), \bar{x}_1) \dots, \bar{x}_1), \bar{x}_1) \text{ for } i \in \{2, \dots, n-1\} \\ E_1 = \frac{g(\bar{x}_n)}{\mu} \\ E_i = \alpha_i \frac{g(\bar{x}_n)}{\mu} \text{ for } i \in \{2, \dots, n-1\} \\ v_0 = E_n f_n(\bar{x}_n), \end{cases} \quad (5.13)$$

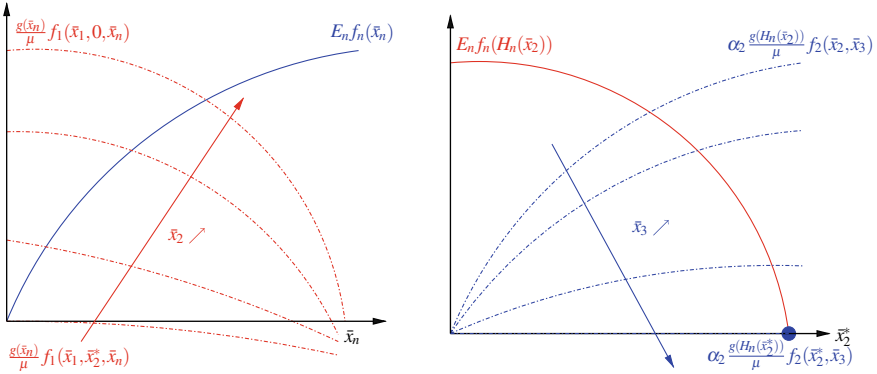
where, for all  $i \in \{2, \dots, n-1\}$ , the functions  $H_i$  are increasing with respect to their arguments and the function  $H_n$  is decreasing (resp. increasing) with respect to its first (resp. second) argument.

*Proof* The proof is inductive.

**Step 1:** Let us first prove that there exists  $x_2^* > 0$  such that, for all  $\bar{x}_2 \in [0, x_2^*]$ , there exists a unique  $\bar{x}_n \geq 0$  such that

$$\frac{g(\bar{x}_n)}{\mu} f_1(\bar{x}_1, \bar{x}_2, \bar{x}_n) = E_n f_n(\bar{x}_n). \quad (5.14)$$

The monotonicity of the functions of the left and the right side of the equation with respect to  $\bar{x}_n$  means that, for all  $\bar{x}_1 > 0$ , there exists  $x_2^* > 0$  such that  $f_1(\bar{x}_1, \bar{x}_2, 0) > 0$



**Fig. 5.3** Intersection of curves  $E_n f_n(\bar{x}_n)$  and  $f_1(\bar{x}_1, \bar{x}_2, \bar{x}_n)$  for all  $\bar{x}_2 \in [0, x_2^*]$  (left) and of curves  $E_n f_n(H_n(\bar{x}_2))$  and  $\alpha_2 \frac{g(H_n(\bar{x}_2))}{\mu} f_2(\bar{x}_2, \bar{x}_3)$  (right)

for all  $\bar{x}_2 \in [0, x_2^*]$  with  $f_1(\bar{x}_1, x_2^*, 0) = 0$ . Then, for all  $\bar{x}_2 \in [0, x_2^*]$ , the left side of Eq. (5.14) is a decreasing function of  $\bar{x}_n$ , is positive for  $\bar{x}_n = 0$  and tends to 0 when  $\bar{x}_n$  goes to infinity. In addition, the right side of the Eq. (5.14) is an increasing function of  $\bar{x}_n$  and is equal to 0 when  $\bar{x}_n = 0$ . Thus, for all  $\bar{x}_2 \in [0, x_2^*]$ , the two curves with respect to  $\bar{x}_n$  necessarily have a unique intersection point. In addition, for  $\bar{x}_2 = x_2^*$ ,  $\bar{x}_n = 0$  is the only solution to the Eq. (5.14) (see Fig. 5.3 left), which concludes the proof of Step 1.

Thus, the function  $\bar{x}_n \triangleq H_n(\bar{x}_2, \bar{x}_1)$  is continuous and decreasing in  $\bar{x}_2$  and can be defined for  $\bar{x}_2 \in [0, x_2^*]$  such that  $H_n(0, \bar{x}_1) > 0$  and  $H_n(x_2^*, \bar{x}_1) = 0$ . For the sake of readability, we omitted the dependence of the equations on  $\bar{x}_1$  in the rest of the proof.

**Step 2:** The rest of the proof is by induction. If the steady-state regime exists, then  $\bar{x}_2$  and  $\bar{x}_3$  are linked by

$$\alpha_2 \frac{g(H_n(\bar{x}_2))}{\mu} f_2(\bar{x}_2, \bar{x}_3) = E_n f_n(H_n(\bar{x}_2)), \tag{5.15}$$

where  $\bar{x}_n$  has been substituted by its expression. As in the first step, we can prove that there exists<sup>2</sup>  $\bar{x}_3^* > 0$  such that, for all  $\bar{x}_3 \in [0, x_3^*]$ , there exists  $x_2 \in [0, x_2^*]$  such that Eq. (5.15) is true. Thus, the function  $\bar{x}_2 \triangleq H_2(\bar{x}_3)$  can be defined, which is well defined, continuous, increasing in  $\bar{x}_3$  for all  $\bar{x}_3 \in [0, x_3^*]$ , and such that  $H_2(0) > 0$  and  $H_2(x_3^*) = x_2^*$  (See Fig. 5.3 right).

**Step 3:** Step 2 is repeated for all  $i \in \{3, \dots, n - 1\}$ . By definition,  $\bar{x}_i$  has to be the solution of the following equation:

<sup>2</sup> In fact,  $x_3^*$  is such that  $f_2(x_2^*, x_3^*) = 0$ , which guarantees that  $f_2(x_2^*, \bar{x}_3) > 0$  for all  $[0, x_3^*]$ .

$$\alpha_i \frac{g(H_n(H_2(\dots(H_{i-1}(\bar{x}_i))))))}{\mu} f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(H_n(H_2(\dots(H_{i-1}(\bar{x}_i))))). \quad (5.16)$$

Then, as in the previous step, it is easy to prove the existence of the function  $H_i$  such that  $x_i \stackrel{\Delta}{=} H_i(x_{i+1})$  is well defined, continuous, increasing in  $\bar{x}_{i+1}$  for all  $\bar{x}_{i+1} \in [0, x_{i+1}^*]$ , and such that  $H_i(0) > 0$  and  $H_i(x_{i+1}^*) = x_i^*$ .

**Step 4:** Through the combination of the results of the previous steps, we can deduce that  $\bar{x}_n$  exists if the following equation has a solution:

$$\bar{x}_n = H_n(H_2(\dots(H_{n-1}(\bar{x}_n)))). \quad (5.17)$$

By definition,  $H_{n-1}$  is defined on  $[0, x_n^*]$  such that  $H_{n-1}(0) > 0$  and  $H_{n-1}(x_n^*) = x_{n-1}^*$ . Let us note that  $H_n(H_2(\dots(H_{n-1}(0)))) > 0$  and  $H_n(H_2(\dots(H_{n-1}(x_n^*)))) = H_n(x_2^*) = 0$ , and because the right side (resp. left side) of Eq. (5.17) is a decreasing (resp. increasing) function in  $\bar{x}_n$ , we can deduce that there exists a unique  $\bar{x}_n \in [0, x_n^*]$ , solution to Eq. (5.17), which concludes the proof.

Remarkably, when all enzymes are reversible, the steady-state regime of the metabolic pathway always exists. We will not develop the results of this structure because the systematic analysis of metabolic pathways indicates the presence of at least one irreversible enzyme per module [17, 20]. In most cases, the irreversible step corresponds to the first or second enzyme. The presence of an irreversible enzyme means that the results of Proposition 1 hold:

**Corollary 4** *Let the assumptions of Proposition 3 be satisfied. If enzyme  $\mathbb{E}_1$  is irreversible, then, for all  $\mu > 0$ ,  $E_n > 0$  and  $\bar{x}_1 > 0$ , there exists a unique steady-state regime  $(\bar{E}_1, \dots, \bar{E}_{n-1})$  and  $(\bar{x}_2, \dots, \bar{x}_n)$  for the system (5.5) such that*

$$\begin{cases} \bar{E}_1 = \frac{g(\bar{x}_n)}{\mu} \\ f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n) \\ v_0 = E_n f_n(\bar{x}_n) \end{cases} \quad (5.18)$$

if and only if  $(\bar{x}_2, \dots, \bar{x}_{n-1})$  exists such that, for  $i \in \{2, \dots, n-1\}$ ,  $\alpha_i \frac{g(\bar{x}_n)}{\mu} f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(\bar{x}_n)$ .

As in Proposition 1, the steady-state regime is only defined by  $f_1$ ,  $g$  and  $f_n$  as long as none of the enzymes saturate. However, the condition of saturation explicitly depends on the steady-state regime and is then less useful.

### 5.3.2 Integration of the Main Biological Configurations

We will now discuss the impact of different biological configurations in more detail. These configurations involve changes due to the presence of isoenzymes, co-factors and co-metabolites. The mathematical results are presented in the most general case of the end-product control structure, in which only the first enzyme is irreversible and the genes are not organised in a single operon.

#### 5.3.2.1 Impact of Co-Metabolites and Co-Factors

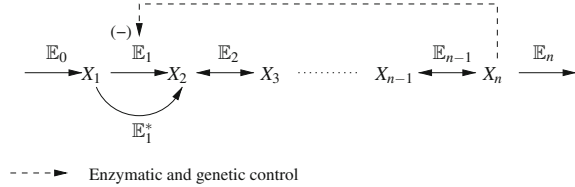
External factors (co-metabolites and co-factors) usually modulate the rate of enzymatic reactions. Co-metabolites, such as ATP/ADP, NAD/NADH or glutamine/glutamate, are also substrates of the enzymes and are transformed into products. Co-factors, such as ions (e.g.  $Mg^{2+}$ ,  $Zn^{2+}$ ) or vitamins, are generally bounded to the enzyme and are therefore considered to be an enzyme component. Both of these types of factors can be easily included in our analysis through the introduction of a new argument in the reaction rate  $f_i$  of the enzyme. Assuming that the  $i$ th reaction requires a co-metabolite, which is labelled as  $X_p$  with concentration  $p$ , then the rate of reaction for an irreversible enzyme (resp. reversible) is given by  $f_i(x_i, p)$  (resp. reversible:  $f_i(x_i, x_{i+1}, p)$ ) such that  $f_i(x_i, 0) = 0$  (resp. reversible:  $f_i(x_i, x_{i+1}, 0) = 0$ ) and the function  $f_i$  is assumed to be increasing in  $p$ .

- 1. The co-metabolite/co-factor acts on the first enzyme,  $\mathbb{E}_1$ .** The maximal flux of the metabolic pathway is given by  $\bar{v}_{n,\max}(\bar{p}) \triangleq \frac{g(0)}{\mu} f_1(\bar{x}_1, 0, \bar{p})$ , where the co-metabolite or the co-factor reaches its steady-state regime  $\bar{p}$ . If the factor decreases the activity of the first enzyme, then  $x_n$  is decreasing and  $E_1$  is increasing.
- 2. The co-metabolite/co-factor acts on the last enzyme,  $\mathbb{E}_n$ .** By definition,  $f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n, \bar{p})$ . The limitation of the concentration of the co-metabolite/co-factor leads to a decrease in the flux demand. Therefore,  $\bar{x}_n$  and  $\bar{E}_1$  are increasing and decreasing functions of  $\bar{p}$ , respectively.
- 3. The co-metabolite/co-factor acts on an intermediate enzyme,  $\mathbb{E}_2, \dots, \mathbb{E}_{n-1}$ .** Remarkably, as long as variations of  $p$  do not lead to enzyme saturation, the steady states of the main components,  $(\bar{v}_n, \bar{x}_n$  and  $\bar{E}_1)$ , remain unchanged.

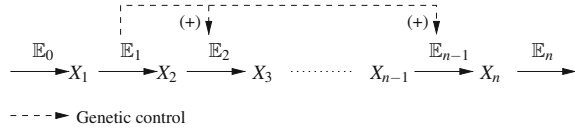
#### 5.3.2.2 Role of an Isoenzyme

Isoenzymes are enzymes that catalyse the same chemical reaction. Let the isoenzyme  $\mathbb{E}_1^*$ , as represented in Fig. 5.4, catalyse the same irreversible reaction as  $\mathbb{E}_1$  (the first reaction).  $\mathbb{E}_1^*$  is not regulated by any intermediate metabolite of the metabolic pathway (neither at the genetic level or at the enzymatic level), which leads to the flux  $E_1^* f_1^*(x_1)$  for  $x_1, E_1^* \geq 0$ .

**Fig. 5.4** Presence of an isoenzyme in the metabolic pathway



**Fig. 5.5** Initial-product control structure



The steady-state regime satisfies the following equation:

$$f_1(\bar{x}_1, \bar{x}_n) \frac{g(\bar{x}_n)}{\mu} + E_1^* f_1^*(\bar{x}_1) = E_n f_n(\bar{x}_n). \quad (5.19)$$

Moreover, the maximal capability of the flux through the metabolic pathway is also modified:

$$\bar{v}_{n,\max} = f_1(\bar{x}_1, 0) \frac{g(0)}{\mu} + E_1^* f_1^*(\bar{x}_1).$$

From Eq. (5.19), as long as the intermediate enzymes do not saturate, the increase of the flux  $\bar{v}_n$  can be obtained either by a decrease in the end product  $\bar{x}_n$  or an increase in the concentration of the isoenzyme  $E_1^*$ .

## 5.4 Other Control Structures

The other elementary module [see Fig. 5.1 (top)], which is named initial-product control structure, usually corresponds to the control structure of degradation pathways. Enzyme synthesis is controlled by the concentration of the first metabolite,  $x_1$ . Due to lack of space, we will only give the condition of existence of the steady-state regime and the qualitative behaviour of the module components to deduce the rules that dictate the connections between modules.

### 5.4.1 Initial-Product Control Structure

We will consider the linear pathway that is shown in Fig. 5.5, which consists of  $n$  metabolites ( $X_1, \dots, X_n$ ) and  $n - 1$  irreversible enzymes ( $E_1, \dots, E_{n-1}$ ) for which the encoding genes are organised in a single operon. The enzyme synthesis is induced



when the concentration of the first metabolite increases. The behaviour of this pathway obeys the following system of differential equations:

$$\begin{cases} \dot{x}_1 = v_0 - E_1 f_1(x_1) \\ \vdots \\ \dot{x}_i = E_1(\alpha_{i-1} f_{i-1}(x_{i-1}) - \alpha_i f_i(x_i)) \\ \vdots \\ \dot{x}_n(t) = \alpha_{n-1} E_1 f_{n-1}(x_{n-1}) - E_n f_n(x_n) \\ \dot{E}_1 = g(x_1) - \mu E_1 \end{cases} \quad (5.20)$$

where  $f_i$  has the same characteristics as in Sect. 5.3 and  $g$  is a positive, continuous and increasing function of  $x_1$  such that for all  $x_1 > 0$ ,  $g(x_1) > 0$ , and  $g(0) = 0$ . We also assume that there exists  $P_{\max} > 0$  such that  $\lim_{x \rightarrow +\infty} g(x) = P_{\max}$ .

**Proposition 4** *For all  $\mu > 0$ ,  $\bar{x}_1 > 0$  and  $E_n > 0$ , there exists a unique steady-state regime  $(\bar{x}_2, \dots, \bar{x}_n)$  and  $(\bar{E}_1, \dots, \bar{E}_{n-1})$  to the system (5.20) such that*

$$\begin{cases} \bar{E}_1 = \frac{g(\bar{x}_1)}{\mu} \\ v_0 = \frac{g(\bar{x}_1)}{\mu} f_1(\bar{x}_1) \\ \bar{x}_i = f_i^{-1} \left( \frac{\alpha_{i-1}}{\alpha_i} f_{i-1}(\bar{x}_{i-1}) \right) \text{ for } i = \{2, \dots, n\} \end{cases} \quad (5.21)$$

if and only if, for all  $i \in \{2, \dots, n-1\}$ ,  $M_1 < \alpha_i M_i$ .

Moreover, the functions  $\bar{x}_i = \bar{x}_i(v_0)$  for  $i = 1, \dots, n$  are increasing in  $v_0$ , the input flux is bounded and the maximal value of  $\bar{x}_1 > 0$  is  $v_{0,\max} \triangleq \frac{P_{\max}}{\mu} M_1$ .

*Proof* The proof of this proposition is straightforward through the writing of the steady-state regime, which, by definition, corresponds to  $v_0 = \frac{g(\bar{x}_1)}{\mu} f_1(\bar{x}_1)$ , and because of the monotonicity of the functions. The existence of the steady-state regime is achieved if and only if the enzymes of the pathway are not saturated. This means that the maximum capacity of each enzyme must be greater than  $v_0$ .

Thus, when  $\bar{x}_1$  is increasing, the flux  $v_0$  and the concentrations of the downstream metabolites are increasing. The IPCS module has also specific properties that can be directly obtained by following the line of the analysis of the EPCS module. The proofs of all of these results are straightforward and are easily deduced from the previous proofs.

#### 5.4.1.1 Comparison Between the Different Control Structures

The two control structures that have been analysed have common characteristics, which were obtained under the assumption that none of the intermediate enzymes saturate:

- the steady-state regime is determined by the characteristics of the first enzyme and its genetic control;
- the maximum capacity of the pathway is limited;
- the co-metabolites of the intermediate enzymes have no impact on the input/output flux or on the genetic control; and
- the presence of an irreversible enzyme prevents the direct spread of the information that is carried by the concentrations of downstream metabolites to the upstream metabolites.

However, there are also notable differences. The EPCS module is inherently driven by the downstream flux demand through  $\bar{x}_n$ , whereas the IPCS module is driven by the upstream flux through  $\bar{x}_1$ . Moreover, the characteristics  $f_n$  of the enzyme  $\mathbb{E}_n$  do not affect the existence of a steady-state regime of the IPCS if the control structure is monotonic. The function  $f_n$  can be increasing or decreasing in  $x_n$ . In other words, a metabolic pathway that is controlled by this type of control structure cannot accommodate a final flux demand of  $v_n$ .

### 5.4.2 Not Controlled Structure

We also introduce a third module, which is named NCS (Not Controlled Structure). This module consists of enzymes that are not genetically or enzymatically controlled by a metabolite in the pathway. The input/output feature of the NCS module at steady state is obtained under the assumption that none of the enzymes of the module saturate and that the first enzyme is irreversible:

$$E_1 f_1(\bar{x}_1) = E_n f_n(\bar{x}_n). \quad (5.22)$$

It follows that the steady-state regime is determined by the concentration of the initial metabolite  $\bar{x}_1$  and by the enzymatic characteristics  $f_1$  and  $f_n$ .

## 5.5 Coordination Between Modules

The mathematical properties that are associated with the two main types of modules have been characterised in the previous sections. We will now discuss the methods by which these modules can be coordinated: global regulations [see Fig. 5.1 (bottom)] and direct connections.

### 5.5.1 Impact of a Global Regulator

In this section, we investigate the impact of a global regulator on the EPCS module. The results for the other structure can be easily deduced. Let us consider that the

synthesis of the first enzyme is also controlled by a global regulator, which leads to

$$\dot{E}_1(t) = g(x_n(t), q(t)) - \mu E_1(t)$$

where  $q(t)$  is the effect of the global regulator. This parameter can also represent any factor that could impact the synthesis of enzyme  $E_1$ .

Assuming that the global regulator reaches its own steady-state regime  $\bar{q}$ , we can deduce from the above results that the global regulator changes the relationship between the concentration of the final product  $\bar{x}_n$ , the flux demand and the enzyme concentration:

$$f_1(\bar{x}_1, \bar{x}_n)g(\bar{x}_n, \bar{q}) = \mu E_n f_n(\bar{x}_n).$$

As long as none of the intermediate enzymes saturate, the global regulator changes the steady-state regime at the level of

- the enzyme concentrations (if the genes are in the same operon):

$$\bar{E}_1 = \frac{g(\bar{x}_n, \bar{q})}{\mu} \quad \text{and} \quad \bar{E}_i = \alpha_i \frac{g(\bar{x}_n, \bar{q})}{\mu} \quad \text{for } i \in \{2, \dots, n-1\},$$

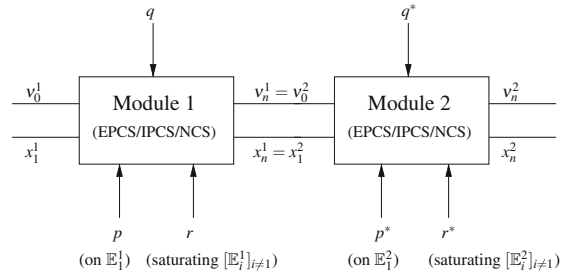
- the end-product concentration:  $f_1(\bar{x}_1, \bar{x}_n)g(\bar{x}_n, \bar{q}) = \mu E_n f_n(\bar{x}_n)$ ,
- the maximal flux capability of the metabolic pathway:  $\bar{v}_{n, \max}(\bar{q}) \triangleq \frac{g(0, \bar{q})}{\mu} f_1(\bar{x}_1, 0)$ ,
- the concentrations of the intermediate metabolites:  $\bar{x}_i = f_i^{-1} \left( \frac{\mu E_n f_n(\bar{x}_n)}{\alpha_i g(\bar{x}_n, \bar{q})} \right)$ .

A global regulator changes the maximum capability of the metabolic pathway directly through the modulation of the concentration of different enzymes in the pathway. Moreover, the flux demand  $v_n$  adapts itself in agreement with the variations induced by the effect of  $q$  on the production function  $g$ .

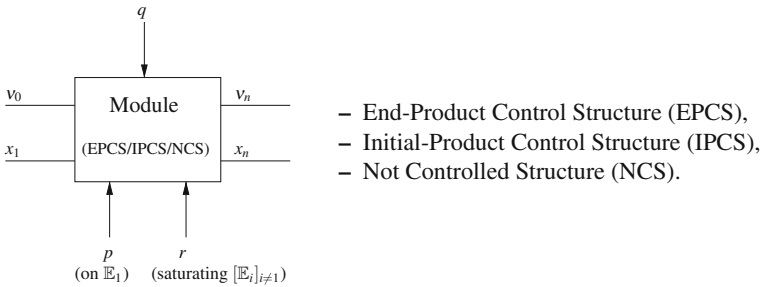
### 5.5.2 Interconnections Between Modules

In this section, we investigate the conditions of existence and uniqueness of a structural steady-state regime for different configurations of connected modules. Therefore, we analysed two modules that are connected in series and in parallel. We will first introduce a generic result for modules that are connected in series and will then provide the rules that define the connection between modules in the summary tables.

**Fig. 5.6** Connection between two modules in series



### 5.5.2.1 The Input/Output Representation of a Module



In steady state, a module is characterised by its input/output characteristics (displayed in the above figure and see Remark 4) whose existence is conditioned by the assumption that the enzymes do not saturate. In the remainder of this section, we assume that this condition of existence is always satisfied. The input/output notations of flux and metabolites are in agreement with systems (5.5), (5.20). We recall the following input/output characteristics, which were obtained for the three types of modules:

- EPCS module:  $\bar{x}_n = H_{pf}(\bar{x}_1)$  and  $v_0 = v_n$ , (the consequences of Corollary 2 are extended for the case of (a) only the first enzyme is irreversible and (b) the genes are not in the same operon), where  $H_{pf}$  is increasing in its argument;
- IPCS module:  $\bar{x}_n = H_{pi}(\bar{x}_1)$  and  $v_0 = v_n$  is defined in Proposition 4, which was extended for the same conditions as the EPCS module, where  $H_{pi}$  is increasing in its argument;
- NCS module:  $\bar{x}_n = H_{ncs}(\bar{x}_1)$  and  $v_0 = v_n$ , where  $H_{ncs}$  is increasing in its argument.

We can deduce the following consequences for two modules that are connected in series (see Fig. 5.6):

- the connection of EPCS modules in series leads to a system with a unique steady-state regime. For the  $i$ th EPCS module, all of the upstream EPCS modules are reduced through the increasing characteristics  $\hat{H}_{pf}$  such that  $\bar{x}_n^i = \hat{H}_{pf}^i(\bar{x}_1^i)$  and  $v_0^1 = v_n^i$  by using  $\bar{x}_1^{k+1} = \bar{x}_n^k$  and  $\bar{x}_n^k = H_{pf}^k(\bar{x}_1^k)$  for  $k \in \{1, \dots, i-1\}$ ;

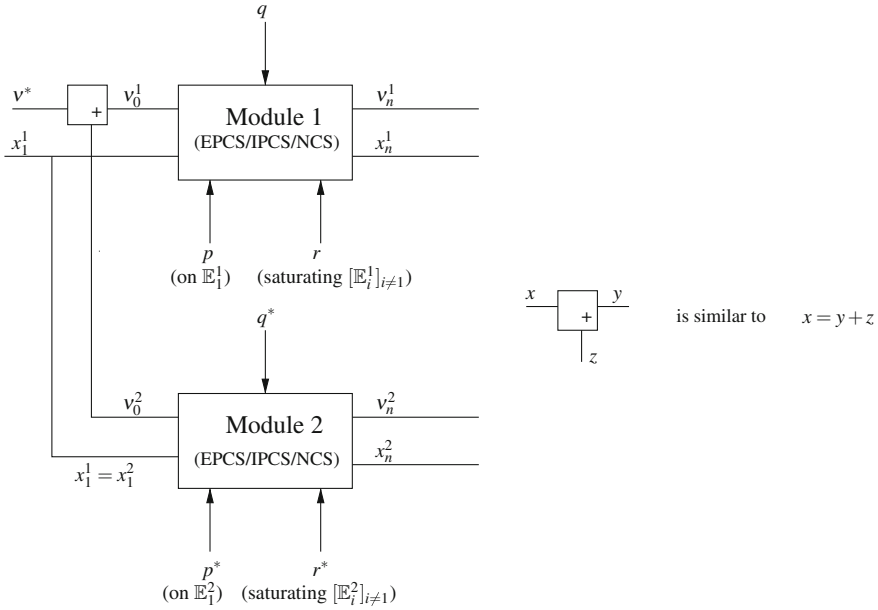


Fig. 5.7 Two modules connected in parallel

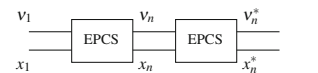
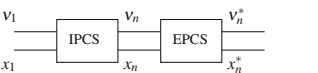
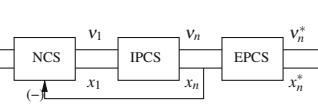
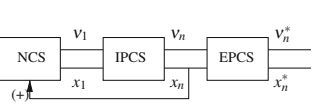
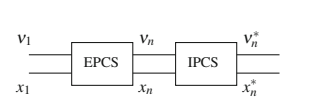
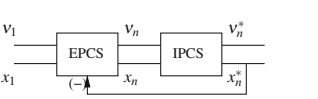
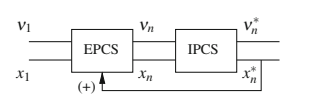
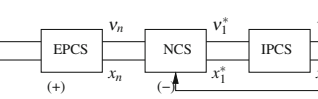
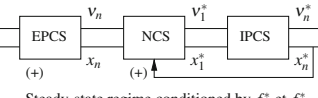
- the connection of IPCS modules in series leads to a system that has a unique steady-state regime. For the  $i$ th IPCS module, all of the upstream IPCS modules are reduced through the increasing characteristics  $\hat{H}_{pi}$  such that  $\bar{x}_n^i = \hat{H}_{pi}(\bar{x}_1^1)$  and  $v_0^1 = v_n^i$  by using  $\bar{x}_1^{k+1} = \bar{x}_n^k$  and  $\bar{x}_n^k = H_{pi}^k(\bar{x}_1^k)$  for  $k \in \{1, \dots, i - 1\}$ .

### 5.5.2.2 The Rules That Define the Connection of Modules

The rules for the interconnection of modules can easily be deduced from the proofs of Propositions 3, 4, 5, which are, respectively, shown in pages 12, 17 and 25, for the connection of modules in series (see Fig. 5.6) or in parallel (see Fig. 5.7) under the assumption that none of the enzymes are saturated. Tables 5.1 and 5.2 summarise the rules of interconnection between modules in series and in parallel, respectively. Specifically, for each of the different connections, these tables show if there exists a structural nonzero steady-state regime and how changes in  $v_1$ ,  $v_n$  and  $v_n^*$  results in variations in  $x_n$ ,  $x_n^*$ ,  $E_1$ ,  $E_1^*$ ,  $x_1$ , and  $x_1^*$ . In both tables, for the sake of readability, we use the following notations:  $f_c$  for increasing functions and  $f_d$  for decreasing functions to describe the monotonicity.

The existence of the equilibrium state is always inferred through the monotonicity of the functions and by assuming a final demand for all last connected modules ( $v_n = E_n f_n(x_n)$ ). In some cases, such as in a connection of NCS/IPCS/EPCS modules in

**Table 5.1** Rules for the interconnection of several modules in series and characteristics of the steady-state regime for variations of  $v_n$ , ( $v_n^*$ ) and  $v_1$  (or  $v_0$ )

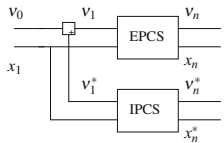
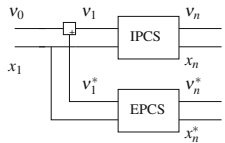
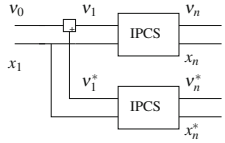
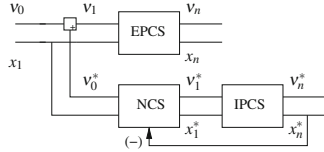
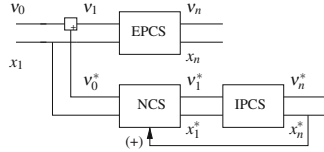
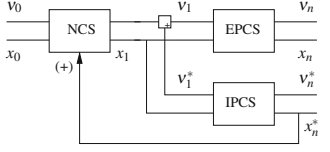
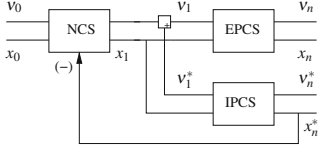
 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_c(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>E_1 = f_c(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>,  <math>x_n = f_c(v_1)</math>, <math>E_1 = f_d(v_1)</math>, <math>x_n^* = f_c(v_1)</math>, <math>E_1^* = f_d(v_1)</math>.</p>	 <p>Existence of nonzero steady-state regime Impossible adaptation of IPCS for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_c(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>,  <math>x_n = f_c(v_1)</math>, <math>E_1 = f_c(v_1)</math>, <math>x_n^* = f_c(v_1)</math>, <math>E_1^* = f_d(v_1)</math>.</p>
 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>x_1 = f_c(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_c(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>x_1 = f_c(v_n^*)</math>, <math>E_1 = f_c(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>,  <math>x_n = f_c(v_0)</math>, <math>E_1 = f_c(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_d(v_0)</math>.</p>	 <p>Steady-state regime conditioned by <math>f_0</math>, <math>f_n</math>, <math>f_n^*</math> Opposite adaptation of IPCS for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>x_1 = f_d(v_n)</math>, <math>E_1 = f_d(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_c(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>x_1 = f_d(v_n^*)</math>, <math>E_1 = f_d(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>,  <math>x_n = f_c(v_0)</math>, <math>x_1 = f_c(v_0)</math>, <math>E_1 = f_c(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_d(v_0)</math>.</p>
 <p>Existence of nonzero steady-state regime Impossible adaptation of IPCS for variations in <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_d(v_n)</math>,  <math>x_n^* = f_d(v_n^*)</math>,  <math>x_n = f_c(v_1)</math>, <math>E_1 = f_d(v_1)</math>, <math>x_n^* = f_c(v_1)</math>, <math>E_1^* = f_c(v_1)</math>.</p>	 <p>No steady-state regime Opposite adaptation of IPCS for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_d(v_n)</math>.</p>
 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_c(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>E_1 = f_c(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>,  <math>x_n = f_c(v_1)</math>, <math>E_1 = f_c(v_1)</math>, <math>x_n^* = f_c(v_1)</math>, <math>E_1^* = f_d(v_1)</math>.</p>	 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>E_1 = f_d(v_n)</math>, <math>x_1^* = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_c(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>E_1 = f_d(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>x_1^* = f_c(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>,  <math>x_n = f_c(v_1)</math>, <math>E_1 = f_d(v_1)</math>, <math>x_n^* = f_c(v_1)</math>, <math>x_1^* = f_d(v_1)</math>, <math>E_1^* = f_d(v_1)</math>.</p>
 <p>Steady-state regime conditioned by <math>f_0^*</math> et <math>f_n^*</math> Impossible adaptation of IPCS for variations in <math>v_n</math>, <math>v_n^*</math>  <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>x_1^* = f_d(v_n)</math>, <math>E_1^* = f_d(v_n)</math>,  <math>x_n = f_d(v_n^*)</math>, <math>E_1 = f_c(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>x_1^* = f_d(v_n^*)</math>, <math>E_1^* = f_d(v_n^*)</math>,  <math>x_n = f_c(v_1)</math>, <math>E_1 = f_d(v_1)</math>, <math>x_n^* = f_c(v_1)</math>, <math>x_1^* = f_d(v_1)</math>, <math>E_1^* = f_d(v_1)</math>.</p>	

We assume that (i) the input flux  $v_1$  (or  $v_0$ ) is able to maintain the concentration of the first metabolite  $x_1$  (or  $x_0$ ) constant and (ii) the enzymes of the modules do not saturate.  $f_c$  increasing function and  $f_d$  decreasing function

series that is associated with positive feedback (see Table 5.1), we cannot directly conclude the existence of a steady-state regime. Typically, we obtain a necessary condition of intersection between two increasing functions:

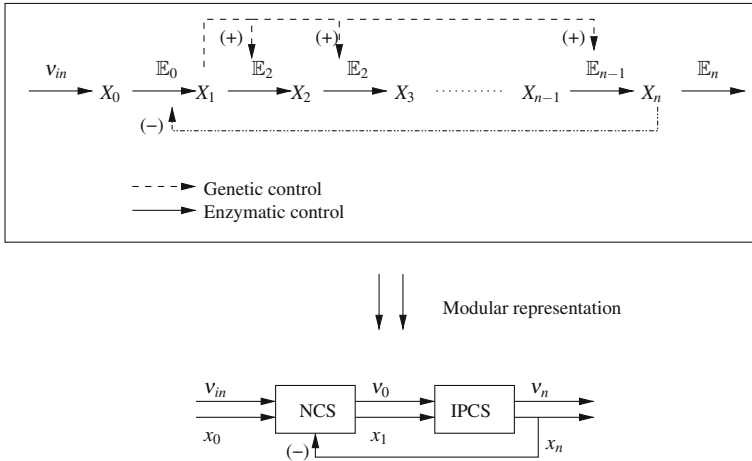
$$E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n),$$

**Table 5.2** Rules for the interconnection of several modules in parallel and characteristics of the steady-state regime for variations of  $v_n$ , ( $v_n^*$ ) and  $v_1$  (or  $v_0$ )

 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math> <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, no effect on <math>v_n^*</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>, no effect on <math>v_n</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_d(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_d(v_0)</math>.</p>	 <p>Existence of nonzero steady-state regime Impossible adaptation of IPCS for variations in <math>v_n</math> <math>x_n = f_d(v_n)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>, no effect on <math>v_n</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_c(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_d(v_0)</math>.</p>
	 <p>Existence of nonzero steady-state regime Impossible adaptation of IPCS for variations in <math>v_n^*</math> <math>x_n = f_d(v_n)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_c(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_c(v_0)</math>.</p>
 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math> <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, no effect on <math>v_n^*</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>x_1^* = f_c(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>, no effect on <math>v_n</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_d(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>x_1^* = f_c(v_0)</math>, <math>E_1^* = f_c(v_0)</math>.</p>	 <p>Steady-state regime conditioned by <math>f_0^+</math>, <math>f_n^+</math> Impossible adaptation of IPCS for variations in <math>v_n^*</math> <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, no effect on <math>v_n^*</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>x_1^* = f_d(v_n^*)</math>, <math>E_1^* = f_d(v_n^*)</math>, no effect on <math>v_n</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_d(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>x_1^* = f_c(v_0)</math>, <math>E_1^* = f_c(v_0)</math>.</p>
 <p>Steady-state regime conditioned by <math>f_0</math>, <math>f_n^+</math> Impossible adaptation of IPCS for variations in <math>v_n</math>, <math>v_n^*</math> <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_1 = f_d(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_d(v_n)</math>, <math>x_n = f_d(v_n^*)</math>, <math>E_1 = f_c(v_n^*)</math>, <math>x_1 = f_d(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_d(v_n^*)</math>, <math>x_1 = f_c(v_0)</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_d(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_c(v_0)</math>.</p>	 <p>Existence of nonzero steady-state regime Feasible adaptation for variations in <math>v_n</math>, <math>v_n^*</math> <math>x_n = f_d(v_n)</math>, <math>E_1 = f_c(v_n)</math>, <math>x_1 = f_d(v_n)</math>, <math>x_n^* = f_d(v_n)</math>, <math>E_1^* = f_d(v_n)</math>, <math>x_n = f_c(v_n^*)</math>, <math>E_1 = f_d(v_n^*)</math>, <math>x_1 = f_c(v_n^*)</math>, <math>x_n^* = f_d(v_n^*)</math>, <math>E_1^* = f_c(v_n^*)</math>, <math>x_1 = f_c(v_0)</math>, <math>x_n = f_c(v_0)</math>, <math>E_1 = f_d(v_0)</math>, <math>x_n^* = f_c(v_0)</math>, <math>E_1^* = f_c(v_0)</math>.</p>

We assume that (i) the input flux  $v_1$  (or  $v_0$ ) is able to maintain the concentration of the first metabolite  $x_1$  (or  $x_0$ ) constant and (ii) the enzymes of the modules do not saturate.  $f_c$  increasing function and  $f_d$  decreasing function

where  $f_0$  and  $f_n$  are both increasing functions of  $\bar{x}_n$ . By convention, the condition of existence of the steady-state regime in these cases is dependent, which is in contrast to those cases in which the existence of the steady state was achieved structurally.



**Fig. 5.8** The modular decomposition of the synthesis of purines: NCS and IPCS modules are connected in series and combined with a negative feedback

### 5.5.2.3 An Example: The Synthesis of Purines

Purines are the main precursors of RNA and DNA synthesis. Thus, one could expect that the control of the synthesis pathway of purines would be driven by the downstream flux demand, i.e., an end-product control structure, such as with amino acids. Surprisingly, the control structure corresponds to an IPCS module that is coupled to an enzymatic inhibition of the upstream enzyme  $E_0$ , which produces the initial metabolite  $X_1$ , by the final metabolite  $X_n$  [27, 30, 39]. We will now prove that, contrary to an IPCS module alone, this control structure is able to cope with a final flux demand. Schematically, the combination corresponds to a NCS module and an IPCS module that are connected in series; these connected modules are combined with negative feedback (see Fig. 5.8). This combination will be referred to as  $IPCS^{(-)}$  in the next section. Moreover, all the genes involved in the purine synthesis are in operon [39].

The steady-state output flux  $v_0$  of the NCS module is given by:

$$v_0 = E_0 f_0(x_0, x_n), \tag{5.23}$$

where  $E_0 > 0$  is fixed and  $f_0$  satisfies the characteristics of an irreversible enzyme that is inhibited by a metabolite and is decreasing (resp. increasing) in  $x_n$  (resp.  $x_0$ ). The flux  $v_0$  is the input flux of module IPCS.

**Proposition 5** *For all  $\mu > 0$ ,  $E_0 > 0$ ,  $E_n > 0$  and  $x_0 > 0$ , there exists a unique steady-state regime  $(\bar{x}_1, \dots, \bar{x}_n)$  and  $(\bar{E}_1, \dots, \bar{E}_{n-1})$  to system (5.20), which is associated with Eq. (5.23), such that*



$$\begin{cases} E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n) \\ v_0 = E_0 f_0(x_0, \bar{x}_n) \\ v_n = E_n f_n(\bar{x}_n) \end{cases} \quad (5.24)$$

if and only if  $v_0 < E_1 M_1$  and for all  $i \in \{2, \dots, n-1\}$ , we have  $v_0 < \alpha_i E_1 M_i$ .

Moreover,

- $\bar{x}_n = \bar{x}_n(x_0)$  is increasing in  $x_0$ .
- $\bar{x}_n = \bar{x}_n(E_n)$  is decreasing in  $E_n$  and  $\bar{x}_i = \bar{x}_i(E_n)$  for  $i = 1, \dots, n-1$  are increasing in  $E_n$ .
- $\bar{x}_i = \bar{x}_i(E_0)$  for  $i = 1, \dots, n$  are increasing in  $E_0$ .

*Proof* The proof is achieved by writing the input/output characteristics of the modules. The connection between the NCS and IPCS in series is direct and the associated characteristics is  $\bar{v}_{n-1} = H_*(\bar{x}_0, \bar{x}_n)$ , where  $H_*$  is increasing in  $\bar{x}_0$  and decreasing in  $\bar{x}_n$ . Then it remains to connect this characteristics with the final flux demand  $E_n f_n(\bar{x}_n) = \bar{v}_n$ , which is increasing in  $\bar{x}_n$ . Due to the monotonicity of the functions  $H_*$  and  $f_n$  with respect to  $\bar{x}_n$ , we conclude the existence and uniqueness of the steady state (under the assumption that the enzymes do not saturate). The behaviour of the module components are deduced from the individual module properties.

Remarkably, the steady-state concentration  $\bar{x}_n$  of the final metabolite is completely determined by the concentrations and the characteristics of the enzymes  $E_0$  and  $E_n$  and not by the enzymes of the IPCS module. For fixed  $E_0$  and  $x_0$ , the input flux  $v_0$  is directly determined as a function of  $v_n$  and  $\bar{x}_n$ . The other components of the IPCS module,  $(\bar{E}_i, \bar{x}_i)$  for  $i \in \{1, \dots, n-1\}$ , are adjusted to cope with the flux demand. In contrast with the case of the IPCS module alone, this module combination is able to cope with the final flux demand.

## 5.6 Decomposition of the Metabolic Network into Modules

### 5.6.1 The Main Identified Combinations

Tables 5.1 and 5.2 show the rules that define the interconnection between modules, regardless of their actual presence in an organism. Using the knowledge-based model of *B. subtilis* [17], we can indicate the actual combination of modules that are present in this organism (and in *E. coli*).

**Connection of EPCS-EPCS modules in series:** This motif, which corresponds to the series of two EPCS modules with an intermediate branching point, occurs in (a) the synthesis of glutamate and glutamine [8, 14, 26, 41, 43], (b) the synthesis of glutamate and proline [7, 8, 26], and (c) the synthesis of S-adenosyl-methionine and cysteine [4, 25]. In *E. coli*, the regulation of the amino acid synthesis pathways have been deeply characterised; therefore, we found that the synthesis of threonine and isoleucine can also be represented by a connected EPCS-EPCS motif [20].

**Connection of IPCS-EPCS modules in series:** We could not identify this type of connection in the metabolic model. However, if we consider that the initial-product control structure is associated with the inhibition of enzyme  $\mathbb{E}_0$  (IPCS<sup>(-)</sup> in Fig. 5.8), the IPCS<sup>(-)</sup>-EPCS connection can be used to represent the connection between the glycolysis pathway and the syntheses of isoleucine, leucine and valine [9, 32, 33, 35, 37].

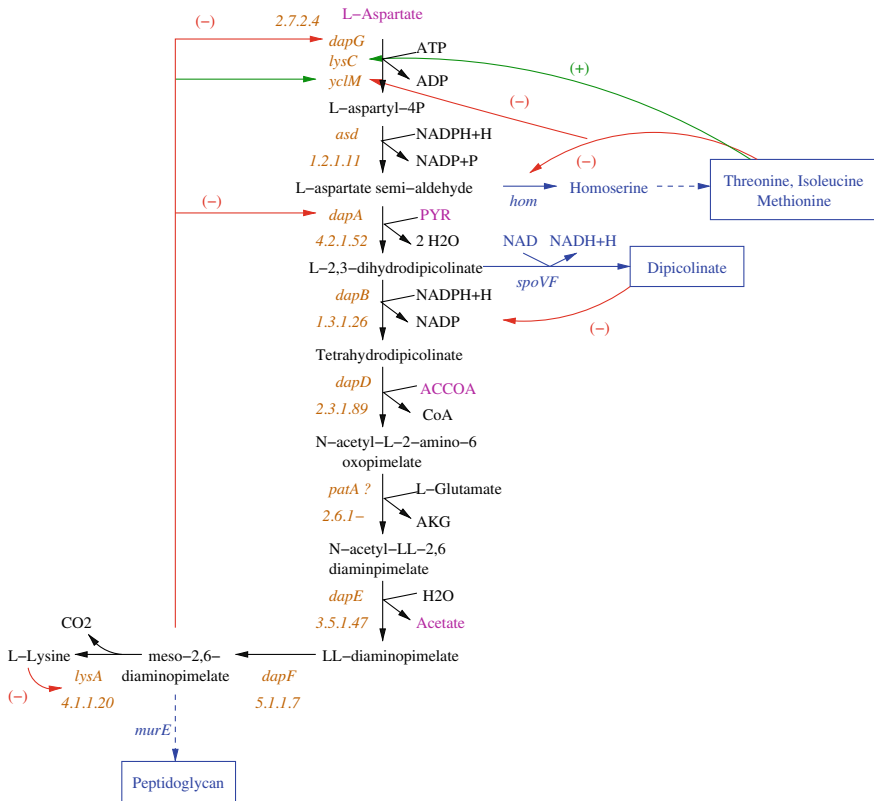
**Connection of EPCS-IPCS modules in series:** The EPCS/IPCS connection is the standard configuration that is used to connect the synthesis and degradation pathways of amino acids, such as arginine [15, 23] and most likely histidine [12, 13, 40, 42]. Unfortunately, the regulation of the synthesis of histidine is unknown. In *E. coli*, the synthesis of histidine is controlled by histidine through the corresponding charged-tRNA and thus by an EPCS module [20]. Usually, a global regulation is present on the connected IPCS module to prevent the simultaneous induction of both modules [6, 12, 13, 42].

**Connection of IPCS-IPCS modules in series:** We identified the presence of the IPCS-IPCS connection at the level of the synthesis and degradation of fatty acids [22, 31]. The global regulator CcpA prevents the degradation of the fatty acids in glycolytic conditions [22, 24]. Moreover, the IPCS/IPCS<sup>(-)</sup> connection connects the degradation of carbohydrates with the glycolysis pathways (see references in [17] and [9]). The IPCS<sup>(-)</sup>-IPCS<sup>(-)</sup> connection has not yet been identified. However, it could exist because the regulatory network is only partially known.

The conditions of existence and uniqueness of the steady-state regime and the qualitative evolution of the main module components can be deduced for all types of these realistic combinations. Remarkably, in most of cases, the steady-state regime exists structurally. Therefore, the existence of steady state only depends on the concentrations of enzymes, which have to be high enough to avoid intermediate enzyme saturation. Finally, the prediction of the qualitative evolution of the main module components has been successfully used to analyse the consistency of datasets (transcriptome, fluxome and metabolome) (see [16] for details).

### 5.6.2 An Example: The Synthesis of Lysine

In this section, we used our results to compare a specific metabolic pathway, the synthesis of lysine, under two distinct physiological conditions: steady-state growth in glucose and in malate. Both of these growth conditions result in similar growth rate values. Therefore, we used two datasets that were produced in the European project BaSysBio (LSHG-CT-2006-037469). Using our approach, we explained the unexpected repression of the lysine pathway that occurs under malate conditions and not in glucose. As will be shown in the rest of the section, this effect is most likely a direct consequence of the high level of aspartate (the first metabolite of the pathway) that is accumulated under malate growth conditions.



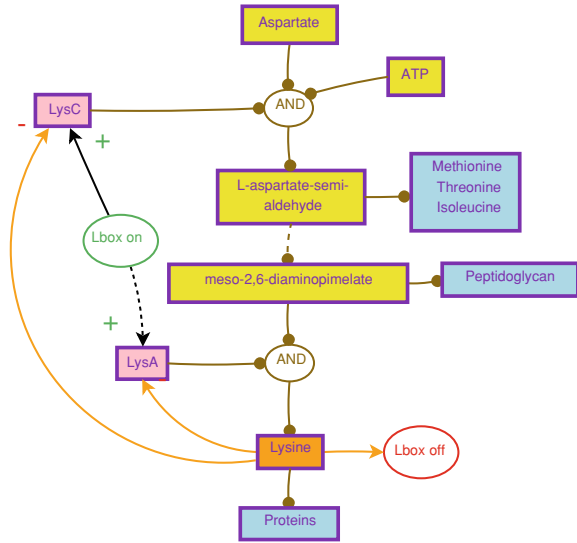
**Fig. 5.9** The synthesis pathway of lysine

Figure 5.9 describes the lysine synthesis pathway and its connections with other essential pathways, whereas Fig. 5.10 highlights the key elements that are involved in the regulation of the lysine synthesis pathway:

- the feedback inhibition of the first enzyme of the pathway, which is encoded by the *lysC* gene, by the end product (lysine) and
- the genetic regulation of the same gene by an L-box mechanism.

The L-box is a RNA riboswitch that involves lysine. Lysine binds directly to the *lysC* nascent mRNA, which causes a structural shift that ends the transcription. The regulation of the *lysA* gene by the same L-box mechanism remains elusive and it is therefore not considered further in the analysis (to maintain the explanation as simple as possible). This structure is classical in metabolic networks and corresponds to the end-product control structure that was described in this chapter. We can directly characterise the properties of the pathway at steady state. The regulation of lysine synthesis satisfies all of the assumptions that are explained in Corollary 4 because

**Fig. 5.10** Regulatory network of lysine synthesis



- LysC is irreversible due to the hydrolysis of ATP,
- The activity of the first enzyme is inhibited by the end product (lysine), and
- The transcription of the first enzyme is repressed by the end-product through an L-box mechanism.

Based on the results described in Sect. 5.3, the expression of the gene *lysC* depends on various factors:

1. metabolites, other than lysine, that act on the first enzyme of the pathway, such as aspartate,
2. flux demand, which is defined mainly by the activity of the tRNA synthase LysS, and
3. external factors that modulate the transcription and translation of the first gene, such as the activity of the RNA polymerases and/or the ribosomes.

The qualitative prediction of the system behaviour with respect to the evolution of the first metabolite (aspartate) and the flux demand (the activity of tRNA synthase LysS) can be predicted (see Table 5.3). The predictions that are shown in Table 5.3 can be extended to any other compatible combinations of conditions. Nevertheless, some qualitative predictions are not possible for some combinations due to their contradictory effects on the system. A contradictory combination, such as an increase in both the flux demand and the aspartate pool, could only be solved if the relative effect of the different factors that act on the regulation is known. Obviously, the knowledge of these factors is related to the identification of system. Because the growth rate between the malate and glucose experiments is similar, the impact of the growth rate on (i) the enzyme synthesis and (ii) the amino acid flux demand by the ribosomes is limited by these two conditions. We thus identified the different

**Table 5.3** Qualitative prediction of the lysine pathway behaviour under various conditions

Considered conditions		Predictions		
LysS ( $E_n$ )	Aspartate	Lysine evolution	Flux evolution	<i>lysC</i> -mRNA
+	Constant	–	+	+
–	Constant	+	–	–
Constant	+	+	+	–
Constant	–	–	–	+

**Table 5.4** Variation of the lysine module components. *gdwc* = gram of cell dry weight

Module components	Glucose	Malate
Aspartate ( $\mu\text{mol/gdwc}$ )	1.4	10.5
Lysine ( $\mu\text{mol/gdwc}$ )	0.1	0.2
mRNA- <i>lysC</i> (log)	14.3	12.3

predictions for a constant flux demand under the two conditions. The concentration of lysine is then an increasing function of the aspartate concentration, and in contrast, the expression of *lysC* is a decreasing function of the aspartate concentration. These predictions are in agreement with the experimental data (see Table 5.4), which led us to conclude that the increasing value of the lysine concentration under malate conditions is most likely due to the increasing aspartate concentration.

## 5.7 Conclusion

The framework that was proposed in this chapter is dedicated to the formal definition and characterization of modules in metabolic pathways. This framework is general enough to study the existence and uniqueness of a structural steady state in any metabolic pathway, including complete metabolic networks. Combined with our results in [17], this is the first report, to the best of our knowledge, of a global-scale analysis of the systematic exploration of all configurations in a realistic biological model. Remarkably, most of the steady-state regimes of realistic metabolic configurations exist structurally. More globally, the local properties of modules have important consequences on the entire metabolic network. Indeed, despite the high coupling that exists in the metabolic pathways (and its associated genetic regulatory network), the steady-state regime of the entire metabolic network is dramatically decoupled. In terms of control, this property is highly expected. Otherwise small variations in a specific module could constantly lead to global genetic adaptations of the entire metabolic network. Beyond the aspects of controllability of the metabolic pathways, we recently shown that the sparing management of resources between the intracellular biological processes of the cell leads to define structural constraints, whose one of their consequences is the emergence of a modular organisation in the metabolic network [18, 19]. An interesting perspective of this framework is the study

of the stability of the elementary modules and their interconnection. The analysis of the stability of metabolic pathways is an open area of research given the very large diversity of configurations and systems and the non-linearity of the equations. Some results have been obtained for linearised systems of specific metabolic pathways [1–3, 36]. Nevertheless, the obtaining of results on the global stability of nonlinear biological system even for one single module remains an open question.

**Acknowledgments** We thank ANR Dynamocell (NT05-2\_44860) and European BaSysBio project (LSHG-CT-2006-037469) for fundings.

## References

1. Alves R, Savageau MA (2000) Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys J* 79:2290–2304
2. Alves R, Savageau MA (2001) Irreversibility in unbranched pathways: preferred positions based on regulatory considerations. *Biophys J* 80:1174–1185
3. Arcak M, Sontag ED (2006) Diagonal stability of a class of cyclic systems and its connection with the secant criterion. *Automatica* 42(9):1531–1537
4. Auger S, Yuen WH, Danchin A, Martin-Verstraete I (2002) The metC operon involved in methionine biosynthesis in *Bacillus subtilis* is controlled by transcription antitermination. *Microbiology* 148(Pt2):507–518
5. Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In: Neidhart FC (ed) *Escherichia coli* and *salmonella*: cellular and molecular biology, 2nd edn. American Society of Microbiology Press, Washington DC, USA, pp 1553–1569
6. Choi SK, Saier MH Jr (2005) Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in *Bacillus subtilis*: potential connection between carbon and nitrogen. *J Bacteriol* 187:6856–6861
7. Chopin A, Biaudet V, Ehrlich D (1998) Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. *Mol Microbiol* 29(2):662
8. Commichau FM, Herzberg C, Tripal P, Valerius O, Stlke J (2007) A regulatory protein-protein interaction governs glutamate biosynthesis in *Bacillus subtilis*: the glutamate dehydrogenase RocG moonlights in controlling the transcription factor GltC. *Mol Microbiol* 65(3):642–654
9. Doan T, Aymerich S (2003) Regulation of the central glycolytic genes in *Bacillus subtilis*: binding of the repressor CggR to its single DNA target sequence is modulated by fructose-1,6-bisphosphate. *Mol Microbiol* 47(6):1709–1721
10. Even S, Pellegrini O, Zig L, Labas V, Vinh J, Brchemmier-Baey D, Putzer H (2005) Ribonucleases J1 and J2: two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E. *Nucleic Acids Res* 33(7):2141–2152
11. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3:121
12. Fisher SH, Rohrer K, Ferson AE (1996) Role of CodY in regulation of the *Bacillus subtilis* hut operon. *J Bacteriol* 178(13):3779–3784
13. Fisher SH, Strauch MA, Atkinson MR, Wray LV Jr (1994) Modulation of *Bacillus subtilis* catabolite repression by transition state regulatory protein AbrB. *J Bacteriol* 176(7):1903–1912
14. Fisher SH, Wray LV Jr (2008) *Bacillus subtilis* glutamine synthetase regulates its own synthesis by acting as a chaperone to stabilize GlnR-DNA complexes. *Proc Natl Acad Sci USA* 105(3):1014–1019

15. Gardan R, Rapoport G, Debarbouille M (1995) Expression of the rocDEF operon involved in arginine catabolism in *Bacillus subtilis*. *J Mol Biol* 249(5):843–856
16. Goelzer A (2010) Emergence de structures modulaires dans les régulations des systèmes biologiques: théorie et applications à *Bacillus subtilis*. PhD thesis, Ecole Centrale de Lyon, Lyon, France. In French
17. Goelzer A, Bekkal Brikci F, Martin-Verstraete I, Noirot P, Bessières P, Aymerich S, Fromion V (2008) Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol* 2:20
18. Goelzer A, Fromion V (2011) Bacterial growth rate reflects a bottleneck in resource allocation. *Biochim Biophys Acta* 1810(10):978–988
19. Goelzer A, Fromion V, Scorletti G (2011) Cell design in bacteria as a convex optimization problem. *Automatica* 47(6):1210–1218
20. Karp PD, Riley M, Saier M, Paulsen IT, Paley SM, Pellegrini-Toole A (2000) The ecocyc and metacyc databases. *Nucleic Acids Res* 28(1):56–59
21. Martinez-Antonio A, Janga SC, Salgado H, Collado-Vides J (2006) Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*. *Trends Microbiol* 14(1):22–27
22. Matsuoka H, Hirooka K, Fujita Y (2007) Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *J Biol Chem* 282(8):5180–5194
23. Miller CM, Baumberg S, Stockley PG (1997) Operator interactions by the *Bacillus subtilis* arginine repressor/activator, AhrC: novel positioning and DNA-mediated assembly of a transcriptional activator at catabolic sites. *Mol Microbiol* 26(1):37–48
24. Miwa Y, Nakata A, Ogiwara A, Yamamoto M, Fujita Y (2000) Evaluation and characterization of catabolite-responsive elements (cre) of *Bacillus subtilis*. *Nucleic Acids Res* 28(5):1206–1210
25. Pelchat M, Lapointe J (1999) In vivo and in vitro processing of the *Bacillus subtilis* transcript coding for glutamyl-tRNA synthetase, serine acetyltransferase, and cysteinyl-tRNA synthetase. *RNA* 5(2):281–289
26. Picossia S, Belitskya BR, Sonenshein AL (2007) Molecular mechanism of the regulation of *Bacillus subtilis* gltAB expression by GltC. *J Mol Biol* 365(5):1298–1313
27. Rappu P, Pullinen T, Mantsala P (2003) In vivo effect of mutations at the prpp binding site of the bacillus subtilis purine repressor. *J Bacteriol* 185(22):6728–6731
28. Santillan M, Mackey MC (2001) Dynamic regulation of the tryptophan operon: a modeling study and comparison with experimental data. *Proc Natl Acad Sci USA* 98(4):1364–1369
29. Sargent MG (1975) Control of cell length. *J Bacteriol* 123(1):7–19
30. Saxild HH, Brunstedt K, Nielsen KI, Jarmer H, Nygaard P (2001) Definition of the *Bacillus subtilis* PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with glyA, guaC, pbuG, xpt-pbuX, yqhZ-foLD, and pbuO. *J Bacteriol* 183(21):6175–6183
31. Schujman GE, Paoletti L, Grossman AD, de Mendoza D (2003) FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Dev Cell* 4(5):663–672
32. Shivers RP, Sonenshein AL (2004) Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol Microbiol* 53(2):599–611
33. Shivers RP, Sonenshein AL (2005) *Bacillus subtilis* ilvB operon: an intersection of global regulons. *Mol Microbiol* 56(6):1549–1559
34. Sontag ED (2002) Asymptotic amplitudes and Cauchy gains: a small-gain principle and an application to inhibitory biological feedback. *Syst Control Lett* 47:167–179
35. Tojo S, Satomura T, Morisaki K, Deutscher J, Hirooka K, Fujita Y (2005) Elaborate transcription regulation of the *Bacillus subtilis* ilv-leu operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA. *Mol Microbiol* 56(6):1560–1573
36. Tyson JJ, Othmer HG (1978) The dynamics of feedback control circuits in biochemical pathways. *J Theor Biol* 5(1):62
37. Ujita S, Kimura K (1982) Fructose-1,6-biphosphate aldolase from *Bacillus subtilis*. *Methods Enzymol* 90(Pt 5):235–241
38. Volkenstein M (1985) Biophysique. Edition Mir

39. Weng M, Nagy PL, Zalkin H (1995) Identification of the *Bacillus subtilis* pur operon repressor. Proc Natl Acad Sci USA 92(16):7455–7459
40. Wray LV Jr, Fisher SH (1994) Analysis of *Bacillus subtilis* hut operon expression indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms: regulation of transcription initiation and inhibition of histidine transport. J Bacteriol 176(17):5466–5473
41. Wray LV Jr, Fisher SH (2005) A feedback-resistant mutant of *Bacillus subtilis* glutamine synthetase with pleiotropic defects in nitrogen-regulated gene expression. J Biol Chem 280(39):33298–33304
42. Wray LV Jr, Pettengill FK, Fisher SH (1994) Catabolite repression of the *Bacillus subtilis* hut operon requires a cis-acting site located downstream of the transcription initiation site. J Bacteriol 176(7):1894–1902
43. Wray LV Jr, Zalieckas JM, Fisher SH (2001) *Bacillus subtilis* glutamine synthetase controls gene expression through a protein-protein interaction with transcription factor TnrA. Cell 107(4):427–435



# Chapter 6

## An Optimal Control Approach to Seizure Detection in Drug-Resistant Epilepsy

Sabato Santaniello, Samuel P. Burns, William S. Anderson and Sridevi V. Sarma

**Abstract** Hidden state transitions are frequent events in complex biological systems like the brain. Accurately detecting these transitions from sequential measurements (e.g., EEG, MER, EMG, etc.) is pivotal in several applications at the interface between engineering and medicine, like neural prosthetics, brain-computer interface, and drug delivery, but the detection methodologies developed thus far generally suffer from a lack of robustness. We recently addressed this problem by developing a Bayesian detection paradigm that combines optimal control and Markov processes. The neural activity is described as a stochastic process generated by a Hidden Markov Model (HMM) and the detection policy minimizes a loss function of both probability of false positives and accuracy (i.e., lag between estimated and actual transition time). The policy results in a time-varying threshold that applies to the *a posteriori* Bayesian probability of state transition and automatically adapts to each newly acquired measurement, based on the evolution of the HMM and the relative loss for false positives and accuracy. An application of the proposed paradigm to the automatic online detection of seizures in drug-resistant epilepsy subjects is here reported.

---

S. Santaniello (✉) · S. P. Burns · S. V. Sarma

Institute for Computational Medicine, Department of Biomedical Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA  
e-mail: ssantan5@jhu.edu

S. P. Burns

e-mail: sburns9@jhu.edu

S. V. Sarma

e-mail: ssarma2@jhu.edu

W. S. Anderson

Department of Neurosurgery, Johns Hopkins School of Medicine,  
600 North Wolfe Street, Baltimore, MD 21287, USA  
e-mail: wanders5@jhmi.edu

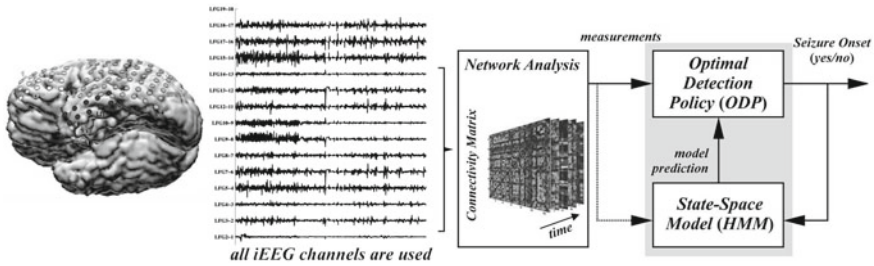
**Keywords** Drug-resistant epilepsy · Seizure onset detection · Hidden Markov model (HMM) · Bayesian estimation · Intracranial electroencephalogram (iEEG) · Optimal control

## 6.1 Hidden State Transition Detection in Medicine

The classic problem of detecting abrupt changes in a sequence of noisy observations collected from a target system [9, 14, 67, 103] has recently gained large interest in neuroscience and medicine (e.g., [1, 7, 39, 44, 56, 61, 84, 91, 94, 96, 97, 102, 104, 109, 110]), mainly because changes in the physiologic expression of a complex biological system (e.g., the brain, heart, liver, etc.) often correspond to critical variations in the clinical or behavioural state. For example, abrupt changes in the atrio-ventricular depolarization delay, the heart rate variability, or the Q–T intervals from the clinical electrocardiogram (ECG) may indicate an incoming tachycardia [19, 50, 96, 97]. The appearance of fast high-frequency oscillations in the intracranial electroencephalogram (iEEG) of an epileptic subject has been recently shown to precede the onset of a seizure and characterize the epileptic focus [4, 27, 43, 107]. Changes in the spiking pattern of somatotopic neurons in the subthalamic nucleus and the globus pallidus can be observed a few hundreds of milliseconds before the actual movement onset of the upper limbs both in non-human primates and Parkinson’s disease patients performing a reach out task [30, 83, 106]. Finally, changes in the spectral content of multi-unit recordings, local field potentials, and iEEG of the primary motor and ventral premotor areas have been shown to encode the target position and the kinematic variables (i.e., arm position and velocity) of reach and grasp movements [8, 28, 51, 55].

In all these examples, the state transitions are hidden in neural measurements and impact statistics computed from these measurements. Detecting timely and accurately such changes would be pivotal to the development of reliable unsupervised monitoring devices, event-based responsive therapies, and more naturally controlled brain-machine interfaces for limb prosthesis. Hence, several methods have been developed in the last twenty years to detect hidden state transitions from sequential neural measurements, and several tools from machine learning, artificial neural network, and estimation theory have been exploited to optimize the detection process [5, 32, 33, 36–38, 44, 45, 54, 68, 73, 91, 93–95, 99]. In particular, it is required that the detection is *online* (i.e., after every newly acquired measurement it must be decided whether or not the change has occurred) and minimizes both the probability of a false positive (i.e., erroneous detection of a state transition) and the lag between actual and estimated change time for true positives [9, 67].

However, the detection algorithms developed thus far often reveal lack of robustness and produce too many false positives when implemented *online* on test data [47], perhaps because none of these methods explicitly introduces performance specifications or a loss function.



**Fig. 6.1** Schematic of the proposed paradigm for seizure onset detection

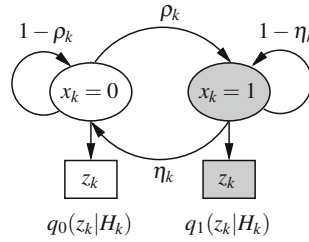
To cope with these issues, we introduced in [79, 80, 82] a Bayesian framework for hidden state transition detection that formulates the change point detection as a “Quickest Detection” (QD) problem [66, 67, 90, 110] and solves it by combining Hidden Markov Models (HMMs) [21, 23], Bayesian Estimation [10], and Optimal Control [11].

In our framework, the brain’s activity is modelled as the output of a two-state HMM where the output neural measurements depend on the actual (not visible) clinical state and are generated sequentially. Based on the HMM evolution and given current and past measurements, we recursively compute the *a posteriori* conditional probability of state transition (Bayesian Estimation), and, finally, we obtain the optimal detection policy (ODP) by minimizing a loss function of the expected distance between actual and detected change time (QD). The loss function is chosen to weight separately early detection (i.e., before the actual change time, which could be a false positive condition) and delayed detection (i.e., after the actual change time), thus penalizing differently the probability of false positives and true positives. This function depends on both the state-transition probability and the sequential measurements, and it is minimized via optimal control [11].

In [75–77] we recently applied this framework to the automatic detection of seizures in drug-resistant epileptic subjects. In this case, we exploited continuous multi-channel iEEG recordings to develop a time-varying spectrum-based matrix of the brain network connectivity and, correspondingly, we computed a measure of the connectivity strength. Then, we assumed that such a measure is the output of a two-state HMM (the states correspond to the seizure and non-seizure condition, respectively) and derived the ODP such that seizure-related changes in the stochastic distribution of this measure were automatically detected. A schematic of our framework for seizure onset detection is reported in Fig. 6.1.

Here, we summarize the main features of our approach and apply it to the seizure onset detection problem on a novel extended dataset of multi-channel iEEG recordings from five drug-resistant epileptic subjects (604 h of recordings, 20 seizures altogether). A specific measure of connectivity strength [76] is also explored in order to better characterize the brain activity at the seizure onset.

Our interest in epilepsy and seizure onset detection is motivated by the fact that (i) epilepsy affects a large population worldwide (over 60 million people) [12, 74,



**Fig. 6.2** HMM schematic with two hidden states ( $x_k = 0$  and  $x_k = 1$ ) and observable output  $z_k$ ,  $k = 1, 2, \dots$ .  $\rho_k$  and  $\eta_k$  are the probabilities of transition from state 0 to state 1 and vice versa, respectively.  $q_x(z_k|H_k)$  is the probability function of  $z_k$  in state  $x \in \{0, 1\}$  conditioned on the past output sequence  $H_k \triangleq \{z_0, z_1, \dots, z_{k-1}\}$

98] and (ii) the drugs currently used to manage the frequency and severity of the seizures are ineffective in over 30 % of the epilepsy patients, with often side-effects due to over-treatment [22, 89]. In this scenario, there is an increasing interest in automated closed-loop intervention approaches (e.g., responsive neurostimulation [3, 25]), but such approaches are most effective when administered immediately prior to or after seizure onset. Therefore, novel unsupervised *online* seizure detection policies are required such that the desired performance measures are tuned non-heuristically by using the design of the loss function.

## 6.2 Hidden Markov Model and State Evolution

We consider the affected brain under the following assumptions: (i) the brain evolves according to a two-state hidden Markov model (HMM), where the state is unknown and “hidden” into a sequence of neural measurements (observations), see Fig. 6.2; (ii) the observations are available at discrete stages  $k = 0, 1, 2, \dots$ ; and (iii) the generic observation  $z_k$  depends on the current state  $x_k$  and the previous observations, which are given in the history sequence  $H_k \triangleq \{z_0, z_1, \dots, z_{k-1}\}$ . Under these hypotheses, the goal of our paradigm is to detect the transition time  $\bar{T} > 0$  from state 0 to state 1.

Note that the formulation, which was derived for a two-state HMM in [75–77, 79, 80, 82], can be generalized for an  $N$ -state HMM, with  $N > 2$ . The HMM is fully characterized by providing the probability of the initial state  $p_0 \triangleq P(x_0 = 1)$ , the conditional probabilities of state transition,  $\rho_k \triangleq P(x_k = 1|x_{k-1} = 0)$  and  $\eta_k \triangleq P(x_k = 0|x_{k-1} = 1)$ , for all  $k$ , and the probability of the output measurement  $q_x(z|H_k) \triangleq P(z_k = z|x_k = x, H_k)$ , for any  $x \in \{0, 1\}$ , history  $H_k$ , and value  $z$  of interest [23]. Finally, note that, because of the use of an HMM and our focus on the transition from state 0 to state 1 only,  $\bar{T}$  is a discrete random variable with  $P(\bar{T} = k) = \rho_k \prod_{j=0}^{k-1} (1 - \rho_j)$ .

In [75, 80] we solved this problem by introducing the Bayesian *a posteriori* probability  $\pi_k$  of being in state 1 at stage  $k$  given the measurements up to and including stage  $k$ , i.e.,  $\pi_k \triangleq P(x_k = 1|z_k, H_k)$ . It can be shown that the variable  $\pi_k$  evolves recursively, based on the current and past observations ( $z_k$  and  $H_k$ , respectively), the probabilities  $\rho_k$  and  $\eta_k$  of state transition in the HMM, and the ratio  $L_k \triangleq \frac{q_1(z_k|H_k)}{q_0(z_k|H_k)}$ :

$$\pi_0 = P(x_0 = 1|z_0) = \frac{q_1(z_0)p_0}{q_0(z_0)(1-p_0) + q_1(z_0)p_0}, \quad (6.1a)$$

$$\begin{aligned} \pi_{k+1} &= \frac{L_{k+1}[\pi_k(1-\eta_{k+1}) + \xi_k\rho_{k+1}]}{(1-\rho_{k+1})\xi_k + \pi_k\eta_{k+1} + L_{k+1}[\pi_k(1-\eta_{k+1}) + \xi_k\rho_{k+1}]}, \quad (6.1b) \\ &\triangleq \Phi_{k+1}(\pi_k, z_{k+1}, H_{k+1}), \end{aligned}$$

where  $\xi_k \triangleq 1 - \pi_k$  [80].

### 6.3 Optimal Control-Based Detection Policy

In [80] we formulated the state transition detection problem as an optimal stopping problem and we stated the required performance goals upfront by minimizing the loss function:

$$J_0 \triangleq a_1 E_{\bar{T}}\{\varphi(\bar{T} - T_S)\}P(\bar{T} > T_S) + a_2 E_{\bar{T}}\{\varphi^2(T_S - \bar{T})\}P(\bar{T} \leq T_S). \quad (6.2)$$

In (6.2),  $\varphi(\varepsilon)$  is a (user-defined) nonnegative and non-decreasing function of the distance ( $\varepsilon$ ) between the estimated and the actual change time ( $T_S$  and  $\bar{T}$ , respectively). We set  $\varepsilon \triangleq \bar{T} - T_S$  or  $\varepsilon \triangleq T_S - \bar{T}$  for early ( $\bar{T} > T_S$ ) or delayed ( $\bar{T} \leq T_S$ ) detection, respectively, and  $\varphi(\varepsilon) = 0$  for  $\varepsilon < 0$ . Furthermore,  $E_{\bar{T}}\{\cdot\}$  in (6.2) is the expected value and parameters  $a_1, a_2 > 0$  are parameters introduced to trade-off between early and delayed detection.

With regard to the specific problem here presented (i.e., seizure onset detection),  $\bar{T} > T_S$  indicates either a false positive (if  $\bar{T}$  does not occur in a reasonably short time window following the warning, see Sect. 6.6) or an early detection (if  $\bar{T}$  does occur within the window). Also, we weighted  $|\bar{T} - T_S|$  differently in case of early and delayed detection (linear vs. quadratic value of  $\varphi$ ) in order to penalize more the occurrence of long delays (i.e.,  $T_S \gg \bar{T}$ ), which are unacceptable in case of responsive treatments to seizures.

We minimized (6.2) by introducing the decision variable  $u_k \in \{0, 1\}$ , which indicates whether a state transition has been detected ( $u_k = 1$ ) or not ( $u_k = 0$ ), and then expanding the original state space model (6.1a, 6.1b):

$$\pi_{k+1} = f_{k+1}(\pi_k, z_{k+1}, H_{k+1}, u_k) \triangleq \begin{cases} \Phi_{k+1}(\pi_k, z_{k+1}, H_{k+1}) & u_k = 0 \\ \text{termination} & u_k = 1 \end{cases} \quad (6.3)$$

In this way, the detection results in deciding the switch time from  $u_k = 0$  to  $u_k = 1$  that minimizes (6.2) [11]. In (6.3) the “*termination*” state indicates that we have stopped caring about the observations  $z_k$ .

The loss function (6.2) can be constructed in terms of  $z_k$ ,  $\pi_k$ , and  $u_k$ . To this purpose, we define a loss-per-stage  $G_k(\pi_k, u_k)$  that penalizes both the missing of a state transition ( $k > \bar{T}$  and  $u_k = 0$ ) and the detection of a transition before its actual occurrence ( $k < \bar{T}$  and  $u_k = 1$ ), while it is 0 otherwise:

$$G_k(\pi_k, u_k) \triangleq \begin{cases} a_2 E_{\bar{T}}\{\varphi(k - \bar{T})\} \pi_k & u_k = 0 \\ a_1 E_{\bar{T}}\{\varphi(\bar{T} - k)\} \xi_k & u_k = 1 \\ 0 & \text{otherwise} \end{cases} \quad (6.4)$$

In (6.4)  $\xi_k \triangleq 1 - \pi_k$  and  $G_k(\pi_k, u_k) = 0$  if the switch from  $u_k = 0$  to  $u_k = 1$  has occurred before the stage  $k$ . We also introduce a terminal loss for missing the state transition over the whole observation horizon  $[0, M)$ :

$$G_M(\pi_M) = \begin{cases} a_1 E_{\bar{T}}\{\varphi(\bar{T} - M)\} \xi_M & u_{M-1} = 0 \\ 0 & \text{otherwise} \end{cases}$$

where  $\xi_M \triangleq 1 - \pi_M$ . In this way, for any policy ( $u_0 = \dots = u_{T_S-1} = 0, u_{T_S} = 1$ ), minimizing the loss function (6.2) corresponds to minimizing the loss function:

$$E_{z_0, z_1, \dots, z_M} \left\{ G_M(\pi_M) + \sum_{k=0}^{M-1} G_k(\pi_k, u_k) \right\} \quad (6.5)$$

provided that  $\varphi$  is non-decreasing. We note that, in the formulation given in [75–77, 79, 80, 82],  $M$  is finite, i.e., the detection problem is restricted to the class of decision policies that stop almost surely (i.e., with probability 1) in finite time, which guarantees that the cumulative loss (6.5) is finite. A generalization to the case  $M \rightarrow \infty$  can be achieved by following [11] (vol. II, chapter 6).

Finally, the minimization of the cost (6.5) is achieved recursively by using Dynamic Programming [11]:

$$J_M(\pi_M) = G_M(\pi_M) \quad (6.6a)$$

$$J_k(\pi_k) = \min \left\{ G_k(\pi_k, u_k = 1), G_k(\pi_k, u_k = 0) \right. \\ \left. + E_{z_{k+1}} \left\{ J_{k+1}(\Phi_{k+1}(\pi_k, z_{k+1}, H_{k+1})) | H_{k+1} \right\} \right\} \quad (6.6b)$$

with  $H_{k+1} \triangleq (H_k, z_k)$ , and the resultant optimal solution (i.e., the estimated state transition time) is

$$T_{ODP} = \min \left\{ 0 < k < M | \pi_k > F_k(\pi_k, z_k, H_k) \right\} \quad (6.7)$$

where

$$F_k(\pi_k, z_k, H_k) \triangleq \frac{a_1 E_{\bar{T}}\{\varphi(\bar{T} - k)\} - \Omega_{k+1}}{a_1 E_{\bar{T}}\{\varphi(\bar{T} - k)\} + a_2 E_{\bar{T}}\{\varphi(k - \bar{T})\}}$$

and

$$\begin{aligned} \Omega_{k+1} &\triangleq E_{z_{k+1}}\{J_{k+1}(\Phi_{k+1}(\pi_k, z_{k+1}, H_{k+1}))|H_{k+1}\} \\ &= \sum_z J_{k+1}(\Phi_{k+1}(\pi_k, z, H_{k+1}))P(z_{k+1} = z|H_{k+1}) \\ &= \sum_z J_{k+1}(\Phi_{k+1}(\pi_k, z, H_{k+1}))\Psi_{k+1} \\ \Psi_{k+1} &\triangleq q_1(z|H_{k+1})(\pi_k + \xi_k p_{k+1}) + q_0(z|H_{k+1})(1 - p_{k+1})\xi_k \end{aligned}$$

and the summation is taken over all the possible values  $z$  of  $z_{k+1}$ .

## 6.4 Network-Based Analysis of the iEEG Recordings

In [75, 77] we proposed to apply the ODP policy (6.7) to the automatic detection of seizure onsets in drug-resistant epileptic subjects. In this case, it is pivotal choosing a sequence of measurements  $z_k$ ,  $k = 0, 1, 2, \dots$ , such that changes in the stochastic distribution of  $z_k$  occur at the onset of the seizures.

Several univariate and bivariate measures have been computed thus far by using single-channel or two-channel EEG signals (both surface and intracranial) [18, 24, 31, 34–36, 40, 41, 49, 52, 58, 59, 61, 69, 73, 93, 99], although none of them has provided a consistent separation between seizure and non-seizure periods. More recently, network-based approaches have been proposed to simultaneously analyse signals from all the available electrodes [6, 16, 46, 65, 71, 72, 86–88, 105]. These approaches treat each electrode as a node in a graph, and any two nodes are connected if the activities at these sites are dependent. The resultant connectivity matrix associated with the graph is analysed and statistics computed from this matrix have revealed significant changes in the graph topology at the seizure onset.

Following this framework, we defined in [75, 77] the connectivity matrix,  $A$ , as the normalized cross-power in a specific epilepsy-related frequency band  $B$  among all the available iEEG signals, i.e., the generic element  $A_{i,j}$  of  $A$  is the cross-power between the iEEG signals recorded by electrode  $i$  and  $j$ . Then, we computed the singular value decomposition of  $A$  and used the leading singular value  $\sigma_1$  as the measurement to be monitored in order to detect a state transition at the seizure onset.

More recently, we noted that the distribution of the whole set of singular values changes over time because of the occurrence of seizures and these changes might be reflected by the median value of the singular values [76]. Therefore, we compute

here  $A$  over a sliding window (length: 2.5 s), with 1 s sliding step (i.e.,  $A$  is updated every second to capture the evolution of the brain topology), and, for each window  $k = 1, 2, \dots$ , we estimate the median value  $\hat{\mu}_{\sigma,k}$  of the singular values of  $A$ . Then, we use the sequence  $\hat{\mu}_{\sigma,k}, k = 1, 2, \dots$  as the output observations  $z_k$  of the HMM in Fig. 6.2. For sake of seizure detection, indeed, we assume that the epileptic brain follows the HMM in Fig. 6.2, with  $x = 1$  ( $x = 0$ ) representing the seizure (non-seizure) condition.

The median  $\hat{\mu}_{\sigma}$  can reflect changes in the distance among the singular values, e.g., it might be very low if there is only one large singular value (i.e.,  $\sigma_1 \gg \sigma_i, i = 2, 3, \dots$ ), while it may become larger in case of a more uniform distribution (i.e.,  $\sigma_i \cong \sigma_j$ , for all  $i, j = 1, 2, 3, \dots$ ), thus reflecting variations of the synchronization level in the brain network [72, 87, 88].

Finally, the computation of  $A$  exploits the frequency band  $B = [13,30]$  Hz for each subject, as the earliest spectral changes around the seizure onset were consistently reported in this band.

### 6.4.1 History-Dependent Model of the Output Measurements

The output probability functions  $q_x(z|H_k), x \in \{0, 1\}$ , in Fig. 6.2 were computed by combining generalized linear models (GLMs) and maximum likelihood estimation [75–77]. Observations were quantized, mapped to integer nonnegative numbers (i.e.,  $n_k \triangleq Q([z_k])$ , with  $n_k \in \mathbb{Z}_0^+$  for all  $k$ ), and fitted by a Poisson law [92]:

$$q_x(z_k = z|H_k) \cong P(n_k = Q([z])|H_k, x) \triangleq e^{-\lambda_{x,k}} \frac{\lambda_{x,k}^{Q([z])}}{Q([z])!} \quad (6.8)$$

where  $\lambda_{x,k}$  is the instantaneous rate and depends on the state  $x$ , time  $k$ , and the previous history  $H_k$ . Then, the time evolution of  $\lambda_{x,k}$  was modeled via GLM [53]

$$\log \lambda_{x,k} = \alpha_x + \sum_{j=1}^L \beta_{x,j} n_{k-j} \quad (6.9)$$

where the parameter vector  $\Theta_x \triangleq \{\alpha_x, \beta_{x,1}, \dots, \beta_{x,L}\}$  is fitted on the data via maximum likelihood estimation [15].

We chose the number of parameters  $L = 10$  in  $\Theta_x$  by minimizing the Akaike's information criterion [2] over a set of candidate models and, for each subject,  $\Theta_x$  was estimated separately for state  $x = 0$  and  $x = 1$  on training data. Training data included 1 h of continuous interictal (i.e., seizure-free) recordings and one seizure period (see Sect. 6.5). The training interictal recordings were collected at least 10 h before any seizure event.



**Table 6.1** Experimental setup

Subject ID	Age/Sex	Seizure origin	Seizure type	# h Recordings	# Seizures	# Electrodes
PT-01	18y/M	P-T	CPS:GTC	134	3	86
PT-02	17y/F	I-O	CPS:GTC	134	2	40
PT-03	49y/F	M-T	CPS	69	4	82
PT-04	14y/M	P-O	CPS:GTC	133	7	77
PT-05	20y/M	L-T	CPS	134	4	108

I-O = inferior occipital lobe; L-T = lateral temporal lobe; M-T = mesial temporal lobe; P-O = parietal occipital lobe; P-T = parietal temporal lobe; CPS = complex partial seizure; CPS:GTC = complex partial seizure with secondary tonic clonic generalization

## 6.5 Multi-channel Intracranial EEG Recordings

The experimental setup includes five drug-resistant epilepsy subjects, who were monitored for approximately one week ( $120.8 \pm 28.96$  h per subject, mean  $\pm$  S.D.) with subdural and depth electrode arrays as part of their pre-surgical evaluation at the Johns Hopkins University Epilepsy Center. The decisions regarding the need for invasive monitoring and the placement of electrode arrays were made independently of our study and solely based on clinical necessity. Acquisition of data for research purposes was done with no impact on the clinical objectives of the patient's stay.

Subjects were implanted with subdural grid arrays, subdural strips, or depth electrode arrays in various combinations as determined by the clinical assessment. Subdural grids have 20–64 contacts per arrays and were used in combination with subdural strips (4–8 contacts) or depths arrays. Intracranial contact locations were documented by post-operative CT co-registered with MRI. Table 6.1 reports subject-specific information, including the number of electrodes, the duration of the intracranial recordings, and the type and origin of the annotated seizures.

Intracranial EEG signals were acquired continuously from each subject by using a Stellate<sup>TM</sup> system (Stellate Systems, Inc., Montreal, QC) with 1000 Hz sampling rate and 300 Hz anti-aliasing filter, converted to EDF format for storage and further processing. Board-certified electroencephalographers (up to three) marked, by consensus, the unequivocal electrographic onset (EO) of each seizure and the period between seizure onset and termination. The seizure onset was marked after visual inspection of the iEEG signals and was indicated by a variety of stereotypical electrographic features, e.g., the early presence of beta-band activity (13–25 Hz), bursts of high frequency oscillations (100–300 Hz), an isolated spike or spike and wave complex followed by rhythmic activity, or an electrodecremental response [60, 70, 85, 107]. These features were typically present in at least one channel at the onset of the seizure. In addition to the inspection of the iEEG recordings, video segments of a video-EEG recordings were analyzed to capture changes in the subject's behavior. For each subject, the performances of the ODP were evaluated based on the distance between the ODP-based detected seizure onsets and the annotated EOs.

The research protocol was reviewed by the Johns Hopkins Institutional Review Board and data was stored in a database compliant with HIPAA (Health Insurance Portability and Accountability Act) regulations.

## 6.6 ODP Performance Evaluation

In order to be consistent with the evaluation criteria used in [36, 45, 57, 61, 73, 99], we measured the performances of the ODP by considering (i) the delay between each estimated seizure onset time and the correspondent EO, (ii) the number of true positive detections (TPs), the number of false positives (FPs), and false negatives (FNs) [75–77]. In particular, we classified each detection as TP or FP if an EO occurred within  $\Delta$  s from the detection time or not, with  $\Delta = 20$  in order to be comparable to [36]. EOs that were not detected were classified as FNs.

Furthermore, we compared the ODP with three widely-used paradigms for change point detection, i.e., the classic Bayesian estimator (BE) [10], the cumulative-sum detector (CUSUM) [9, 62, 96, 97], and the threshold-based detector (HT), where the threshold is fixed *a priori* to an heuristically-chosen value. The estimated seizure onset given by BE, CUSUM, and HT is:

- *BE*:  $T_{BE} \triangleq \min\{0 < k < M | \pi_k > 0.5\}$ ,
- *CUSUM*:  $T_{CU} \triangleq \min\{0 < k < M | g_k > \bar{g}\}$ ,
- *HT*:  $T_{HT} \triangleq \min\{0 < k < M | z_k > \bar{h}\}$ ,

with  $\bar{g} \triangleq \mu_g$  and  $\bar{h} \triangleq \hat{\mu}_z$ , respectively, where  $\mu_g$  is the mean value of the CUSUM variable  $g_k$  and  $\hat{\mu}_z$  is the median of the output sequence  $\hat{\mu}_{\sigma,k}$  during the first seizure period (training data), respectively. Note that the CUSUM variable  $g_k$  is defined as:

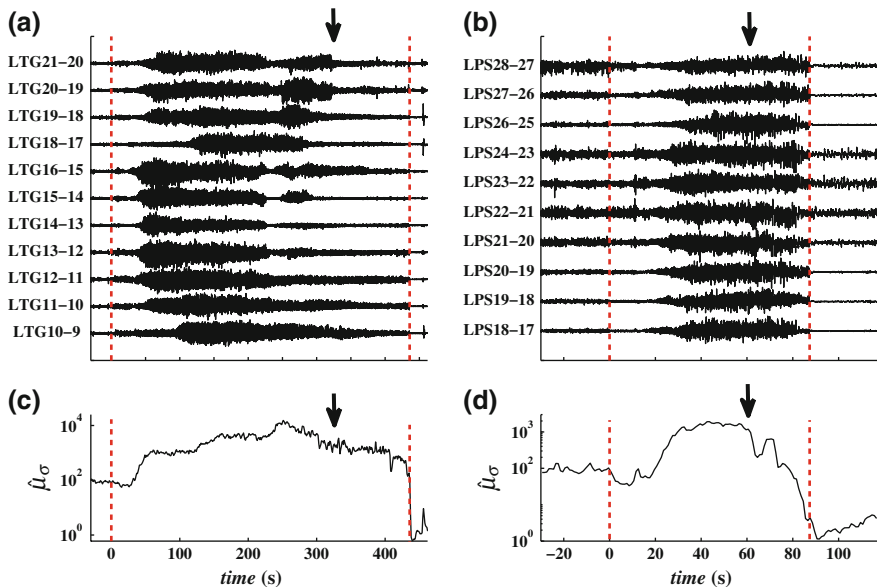
$$g_0 \triangleq 0$$

$$g_k \triangleq \begin{cases} g_{k-1} + l_k & \text{if } g_{k-1} + l_k > 0 \\ 0 & \text{otherwise} \end{cases}$$

with  $l_k \triangleq \ln\left(\frac{q_1(z_k|H_k)}{q_0(z_k|H_k)}\right)$  computed at each stage  $k$  [9, 62].

## 6.7 Results

The ODP framework was used for seizure detection in five drug-resistant epileptic subjects. One hour of seizure-free multi-channel iEEG recordings and one hand-annotated seizure per subject were used to estimate the parameters of the model (6.9) in state  $x = 0$  and  $x = 1$ , respectively, while the remaining data was used to



**Fig. 6.3** Examples of complex partial seizure with (a, c) and without (b, d) tonic-clonic generalization. **a, b** iEEG recordings from multiple focal electrodes (differential montage, labels on the y-axis). **c, d** Median of the singular values for the iEEG recordings in **a** and **b**, respectively. Dashed red lines indicate the hand-annotated EO and termination of each seizure, arrows indicate the onset of the post-convulsion phase. Plots **a, c** refers to subject PT-01 (seizure #1), plots **b, d** refers to subject PT-03 (seizure #1). Log-scale in **c, d** emphasizes the dynamics around the seizure onset.

validate the detection policy. The state transition probabilities  $\rho_k$  and  $\eta_k$  in (6.1b) were assumed time-invariant and estimated for each subject via maximum likelihood [21, 23]. For sake of simplicity, we implemented the policy (6.5)–(6.7) with the linear penalty  $\varphi(\varepsilon) = 2\varepsilon - 1$ , which was introduced in [75–77, 80]. Results are reported in Figs. 6.3, 6.4, 6.5, 6.6, 6.7 and Tables 6.2, 6.3.

### 6.7.1 Network-Based Connectivity Matrix and Singular Value Decomposition

Figure 6.3 reports the median  $\hat{\mu}_\sigma$  of the singular values around the onset and termination of a complex partial seizure (CPS), both with and without secondary tonic-clonic generalization (GTC).

CPS and GTC seizures differ for the duration of the event (GTC seizures are usually longer than CPSs) and the extension of the brain region involved (GTC seizures involve a larger part of the brain, frequently the whole brain), and may be elicited by different pathologic mechanisms [17, 26, 41]. In particular, GTC seizures

are more severe than simple CPSs, as they quickly spread from a small area (i.e., the "focus") to a wide region of the brain, often causing loss of consciousness and convulsions.

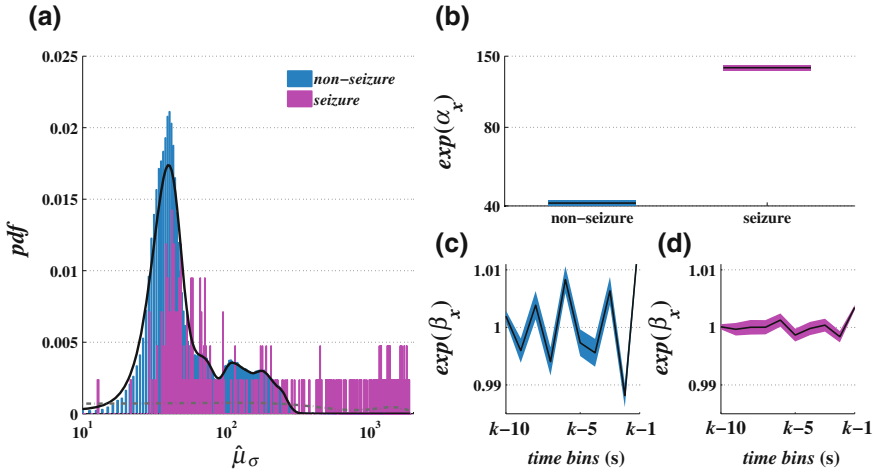
Despite these physiologic and clinical differences, however,  $\hat{\mu}_\sigma$  has similar dynamics in CPS and GTC seizures (Fig. 6.3c, d) and captures a seizure-related increment of the oscillatory activity of the iEEG signals in the band [13, 30] Hz (Fig. 6.3a, b). In particular,  $\hat{\mu}_\sigma$  is stable and shows minor fluctuations both before and after seizure, which correspond to steady-state condition. At the seizure onset, instead,  $\hat{\mu}_\sigma$  shows a recurrent pattern, which is consistent across the subjects and types of seizure: it slowly decreases *first* and *then* rapidly increases, thus reaching a local maximum approximately half of the seizure event. Finally,  $\hat{\mu}_\sigma$  decreases to very small values (i.e., smaller than before the seizure onset) before rapidly returning to the steady state conditions.

The duration of the second decreasing phase of  $\hat{\mu}_\sigma$  (i.e., the one right before the return to steady-state conditions) depends on the extension of the post-convulsion phase, which is the relaxation following a paroxysmal activity (arrows in Fig. 6.3a, b) and usually lasts longer in GTC seizure than simple CPS. Furthermore, the mean value of  $\hat{\mu}_\sigma$  at the seizure termination is lower than the pre-seizure value and, in case of GTC seizures, it shows an abrupt drop, eventually followed by oscillations before returning to steady state conditions (Fig. 6.3c). These results are consistent across all the subjects, despite the various origins and types of the annotated seizures. In particular, the pattern of  $\hat{\mu}_\sigma$  captures the sequence of changes that occur in brain complexity at the transition from non-seizure to seizure activity and reflects an overall increase and more uniform distribution of the singular values 20–50 s after the seizure EO. This pattern might be due to an initial desynchronization at the seizure onset and then a subsequent strong re-synchronization across different brain regions, as suggested by the correlation analysis [87, 88]. Furthermore, (i) the low in value of  $\hat{\mu}_\sigma$  at the seizure termination, (ii) the subsequent long drift toward the pre-seizure steady-state values, and (iii) the eventual post-seizure oscillations (GTC seizures) could overall denote a post-seizure reset of the brain dynamics, with final desynchronization among the different brain regions and lower iEEG activity [41].

## 6.7.2 History-Dependent Output Distributions

Figure 6.4 shows the estimated parameters of model (6.9) both in seizure and non-seizure conditions for the subject reported in Fig. 6.3b, d. A history-independent kernel-smoothing estimation [13] of the probability distribution function of  $\hat{\mu}_\sigma$  is reported in Fig. 6.4a.

It can be noted that the mean value and the variance of  $\hat{\mu}_\sigma$  increase at the seizure onset but there is a large overlap between the history-independent probability distribution functions in seizure and non-seizure conditions (Fig. 6.4a), which determines a poor estimation of the Bayesian *a posteriori* probability  $\pi_k$ , see Fig. 6.5a. In particular, the history-independent estimations of  $q_0(\cdot)$  and  $q_1(\cdot)$  have small amplitude

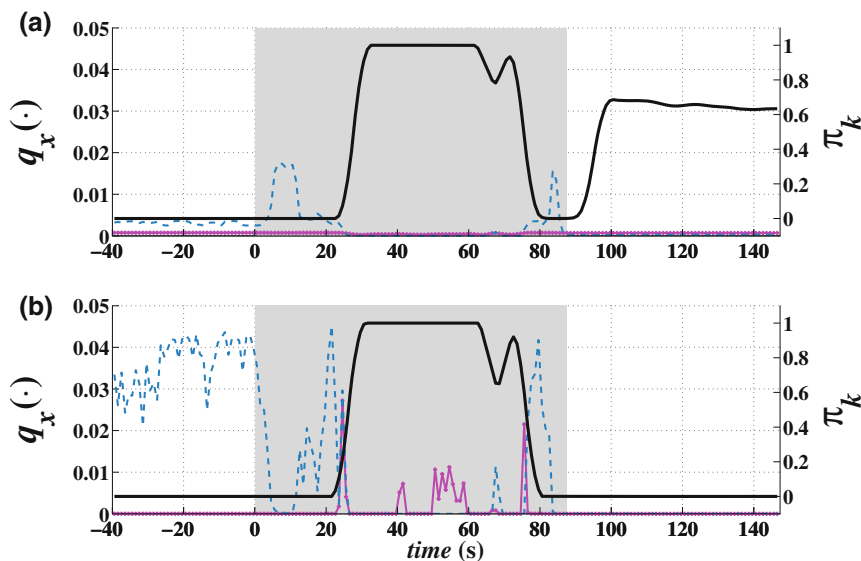


**Fig. 6.4** History-independent versus history-dependent probability of the output observations. **a** Sample distribution of  $\hat{\mu}_\sigma$  in non-seizure (blue bars) and seizure (purple bars) conditions. A history-independent kernel smoothing density estimation of the probability density function (pdf) is given for seizure (gray dash-dotted line) and non-seizure (black line) periods. **b–d** History-dependent parameters of model (6.9) and 95 % confidence bounds (strips) during non-seizure (**b, c, blue**) and seizure (**b, d, purple**) periods. Data refers to subject PT-03

and poorly modulate at the seizure onset (mostly  $q_0$ ) while, at the termination of the seizure, both  $q_0$  and  $q_1$  are  $\sim 0$  because of the post-ictal resetting phase. As a consequence, the probability  $\pi_k$  does not entirely follow the modulation of  $\hat{\mu}_\sigma$ , i.e., it correctly rises from 0 to 1 during seizure (Fig. 6.3d and Fig. 6.5a), but then it has an erroneous late increment to values above 0.5 (the chance level) during the post-ictal phase, which depends on the poor post-ictal modulation of  $q_0$  and  $q_1$ . According to (6.1b), the condition  $\pi_k > 0.5$  erroneously indicates that state  $x = 1$  (seizure) is more likely than state  $x = 0$  (non-seizure) and might lead to a false positive detection of state transition.

The history-dependent model (6.8) and (6.9), instead, indicates that, at any given time  $k$ , the probability of the current value of  $\hat{\mu}_\sigma$  depends on the pattern in the previous 10 s and that such dependency actually varies with the current state  $x = 0$  or  $x = 1$  (Fig. 6.4c, d). In particular, the maximum likelihood estimation of parameter  $\alpha_x$  in (6.9) significantly increases in seizure conditions (Fig. 6.4b) as a consequence of the higher average value of  $\hat{\mu}_\sigma$  during seizure, while the parameters  $\beta_{x,j}$ ,  $j = 1, \dots, 10$  show larger oscillations and (slightly) larger 95 % confidence bounds in case of non-seizure data (Fig. 6.4c), thus indicating higher variability (i.e., larger differences) across consecutive observations and recurrent periodic fluctuations in the sequence  $\hat{\mu}_{\sigma,k}$ .

Differences in model parameters contribute to fit the GLM structure (6.9) to the actual data sequences and allow a selective modulation of probabilities  $q_0(\cdot)$  and  $q_1(\cdot)$  (Fig. 6.5b). In particular,  $q_1$  selectively increases during the seizure, it reaches



**Fig. 6.5** Bayesian probability  $\pi_k$  of state transition (black line) and output probabilities  $q_0(\cdot)$  (blue dashed line) and  $q_1(\cdot)$  (purple line) estimated around seizure #1 in subject PT-03 (see Fig. 6.3b, d). **a** Probabilities obtained by using the history-independent kernel smoothing density estimation in Fig. 6.4a. **b** Probabilities obtained by using the history-dependent GLM (6.9) with parameters reported in Fig. 6.4b–d. Scale on the left and right y-axis in (a, b) refers to  $q_x(\cdot)$ ,  $x = 0, 1$ , and  $\pi_k$ , respectively

an initial peak a few seconds after the hand-annotated EO, and, then, it reaches a final peak approximately at the beginning of the convulsive phase (40–60 s after the EO, see Fig. 6.3b, d), while it is generally low for non-seizure data sequences. Vice versa,  $q_0$  is much higher than  $q_1$  on seizure-free data sequences and shows slow fluctuations, while it quickly decreases to approximately 0 a few seconds after the hand-annotated EO. The opposite dynamics of  $q_0$  and  $q_1$  (i) triggers  $\pi_k$  from 0 to 1 just a few seconds after the seizure onset, (ii) causes a fast decrease of  $\pi_k$  toward the end of the seizure, and (iii) keeps otherwise  $\pi_k$  very low during non-seizure periods.

Interestingly, there are two peaks in the value of  $q_0$  during the seizure, one approximately 20 s after the onset and one toward the end of the seizure (Fig. 6.5b). The first peak accounts for the initial decrease of  $\hat{\mu}_{\sigma,k}$  at the seizure onset, while the last one corresponds to the drop in the value of  $\hat{\mu}_{\sigma,k}$  at the beginning of the post-seizure phase (see Fig. 6.3d).

**Table 6.2** Performance analysis on validation data

Subject ID	ODP			BE			HT			CUSUM		
	FPR	TPR	FN	FPR	TPR	FN	FPR	TPR	FN	FPR	TPR	FN
	(FP/h)	(%)		(FP/h)	(%)		(FP/h)	(%)		(FP/h)	(%)	
PT-01	<b>0.07</b>	<b>100</b>	<b>0</b>	0.04	100	0	0.01	100	0	0.04	100	0
PT-02	<b>0.29</b>	<b>100</b>	<b>0</b>	0.22	100	0	0.14	100	0	0.27	100	0
PT-03	<b>0.07</b>	<b>100</b>	<b>0</b>	0.06	100	0	0.04	100	0	0.03	100	0
PT-04	<b>0.01</b>	<b>100</b>	<b>0</b>	0.01	100	0	0.01	100	0	0.02	100	0
PT-05	<b>0.04</b>	<b>100</b>	<b>0</b>	0.04	100	0	0.05	100	0	0.05	100	0

### 6.7.3 Optimal Detection Policy for Seizure Events

The performances of the proposed optimal policy (6.7) and of the other detectors in Sect. 6.6 are reported in Tables 6.2 and 6.3. Results were obtained by using validation data only (599 h of continuous iEEG recordings including 15 annotated seizures).

All the policies show high sensitivity (100 % of TPs, i.e., all the seizures were correctly detected, Table 6.2), low false positive rates (FPR always lower than 0.29 FP/h, which is required for potential clinical applications), and high concentration of the (eventual) FPs in a small time window around the actual seizures.

The fact that the false positive rate (FPR) is low also with BE and HT suggests that the median  $\hat{\mu}_\sigma$  selectively increases only during the ictal events and therefore provides a robust feature to accurately separate the seizures from the remaining activity. In particular, the pattern of  $\hat{\mu}_\sigma$  (Fig. 6.3c, d) and the high values achieved at the onset of the paroxysmal phase of each seizure (20–50 s after the hand-annotated EOs, Fig. 6.3a, b) account for the low FPR achieved with the HT detector. In this case, however, the choice of the threshold  $\bar{h}$  is pivotal to avoid interictal spikes and outliers in the sequence  $\hat{\mu}_{\sigma,k}$ . We chose of  $\bar{h} \triangleq \hat{\mu}_z$  retrospectively in order to detect only the high values that occur at the beginning of the paroxysmal phase and have low sample probability during the non-seizure periods (Fig. 6.4a). However, despite a high specificity value (average FPR:  $0.05 \pm 0.053$ ), this choice of  $\bar{h}$  determined large detection delays (Table 6.3) and required approximately 40 % of each ictal period ( $41.87 \pm 10.38$  %) to detect the ongoing seizure, which is generally too much for clinical applications.

Smaller delays were achieved with the BE and CUSUM detectors (Table 6.3) and a slightly smaller portion of each ictal period was required to detect the seizures ( $35.99 \pm 11.56$  % and  $40.97 \pm 17.73$  % for BE and CUSUM, respectively) but the FPR significantly increased (average:  $0.07 \pm 0.08$  and  $0.08 \pm 0.11$  for BE and CUSUM respectively), mostly because the Bayesian probability  $\pi_k$  and the cumulative sum variable  $g_k$  increase with the median  $\hat{\mu}_\sigma$  but with a faster pace, i.e., they have steeper slope than  $\hat{\mu}_\sigma$  (Fig. 6.3d and Fig. 6.5b).

The ODP, instead, addressed the trade-off between specificity (i.e., low FPR) and detection delay through the design of a suitable the loss function (6.2). In particular,

**Table 6.3** Detection delay on validation data

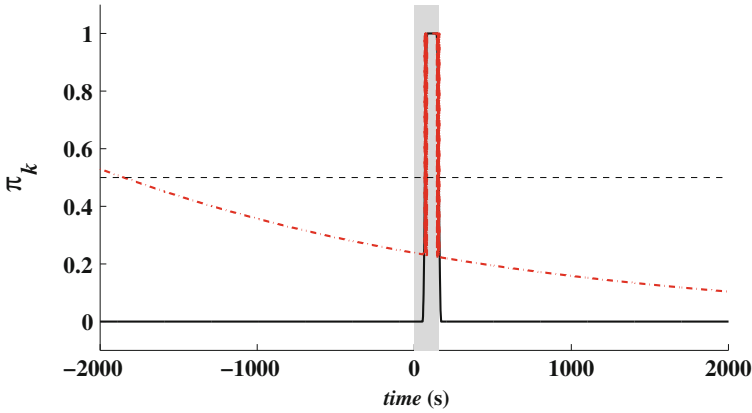
Subject ID	Seizure		Detection delay (s)			
	label #	duration (s)	ODP	BE	HT	CUSUM
PT-01	2	236.4	<b>35.79</b>	48.79	76.79	51.79
	3	358.8	<b>43.20</b>	57.20	135.20	60.20
PT-02	2	82.99	<b>18.06</b>	27.06	34.06	28.06
PT-03	2	77.74	<b>10.33</b>	24.33	26.33	26.33
	3	82.23	<b>14.25</b>	28.25	29.25	29.25
	4	86.43	<b>20.42</b>	33.42	34.42	79.42
PT-04	2	188.2	<b>85.07</b>	98.07	105.07	99.07
	3	203.9	<b>88.45</b>	101.45	105.45	102.45
	4	260.0	<b>95.45</b>	109.45	124.45	110.45
	5	290.1	<b>98.82</b>	112.82	127.82	114.82
	6	366.2	<b>124.20</b>	138.20	147.20	140.20
	7	377.9	<b>180.34</b>	194.34	215.34	196.34
	mean	195.5	<b>58.89</b>	71.62	85.36	76.76
S.D.	112.9	<b>50.90</b>	51.59	57.51	50.02	

we used a ratio  $a_2/a_1 = 1000$  and set  $M \triangleq 1/\rho$  (i.e., reciprocal of the probability of transition from state 0 to state 1, Fig. 6.2) in order to penalize more the detection delay, and we finally achieved a significantly lower delay (paired-sample  $t$ -test,  $p < 0.0005$  Bonferroni corrected, see Table 6.3) and required a significantly smaller portion of each ictal period to detect the seizures ( $27.26 \pm 13.48$  %, paired-sample  $t$ -test,  $p < 0.01$  Bonferroni corrected), which may be feasible for clinical applications, while the increment of the FPR was still limited (average:  $0.09 \pm 0.11$ ) and compatible with potential clinical applications.

Figure 6.6 shows the dynamics of the threshold  $F_k(\pi_k, z_k, H_k)$  in (6.7) around a seizure event. Both before and after the seizure, the low values of the probability  $\pi_k$  (i.e.,  $\sim 0$ ) result in a monotonically decreasing threshold. This is a consequence of the problem formulation (6.2)–(6.7) with a finite  $M$  and the evolution model (6.1a, 6.1b), and reflects the important fact that the likelihood of a seizure increases over time when no seizure is detected. In this case, the choice of  $M$  guarantees a sufficiently long evolution window (the optimal policy is reset and restarts every  $M$  samples or right after a detection) which is related to an estimation of the average inter-time between consecutive seizures.

At the seizure onset, instead, the probability  $\pi_k$  begins to increase (zoom, Fig. 6.7, bottom row) and such change in dynamics is captured by the non-monotonic behaviour of the threshold  $F_k(\cdot)$ . In particular,  $F_k(\cdot)$  has an initial abrupt increment, which is aimed at avoiding potential outliers in the value of  $\pi_k$  and, then, it remains constantly high (i.e.,  $\sim 1$ ) as a consequence of the steady state value  $\pi_k = 1$  during the seizure.





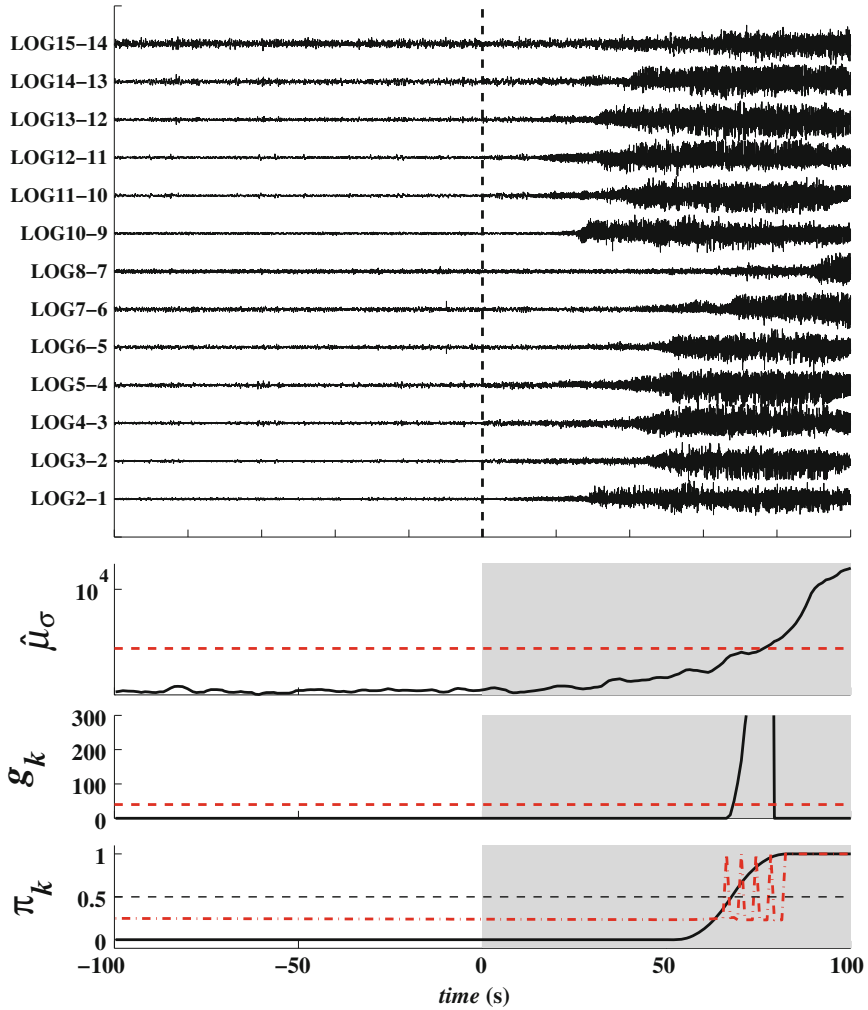
**Fig. 6.6** Probability  $\pi_k$  of state transition (black line) and ODP threshold  $F_k(\pi_k, z_k, H_k)$  (red dash-dotted line) estimated around a seizure (gray background). Data refers to seizure #1 of subject PT-04

The advantage of the adaptive threshold  $F_k(\cdot)$  over the fixed thresholds used with the BE, HT, and CUSUM detectors is reported in Fig. 6.7, where the fast decreasing dynamics of  $F_k(\cdot)$  allows to capture an early consistent modulation of the probability  $\pi_k$  well before it reaches the BE threshold value of 0.5. Also, the dynamics of the sequences  $\hat{\mu}_{\sigma,k}$  and  $g_k$  is slow during the first part of the seizure and more than 30 % of the ictal period is required to detect a noticeable change in these sequences, which may account for the long delays reported by the HT and CUSUM detectors. The Bayesian probability  $\pi_k$ , instead, reveals an earlier modulation, which is determined by the changes captured with the estimated model parameters  $\Theta_x, x \in \{0,1\}$  in (6.9).

## 6.8 Discussion

We recently developed an optimal control-based framework for change point detection and we used it to automatically detect seizure events in drug-resistant epilepsy subjects [75–77, 79, 80, 82]. In particular, we model the evolution of the affected brain as a hidden Markov chain [23], track the Bayesian probability of a state transition [10], and finally detect the seizure onset by solving a Quickest Detection problem [66, 67, 90, 110] via Dynamic Programming [11].

As noted in [80], this framework generalizes the well-known problem of online change-point detection [9, 14, 67, 103] to a class of output measurements which are non-binary and history-dependent, thus resulting of interest for applications in neuroscience and medicine. Also, the optimization problem does not require a specific type of probability distribution for the change times  $\bar{T}$  (which follows from the



**Fig. 6.7** Zoom in around the seizure reported in Fig. 6.6. From *top to bottom*: iEEG signals recorded across multiple focal electrodes (differential montage, labels on the y-axis) around the hand-annotated EO (*dashed vertical line*); median  $\hat{\mu}_\sigma$  of the singular values estimated from the iEEG signals above (*black line*) and HT threshold  $\bar{h}$  (*red dashed line*); CUSUM variable  $g_k$  correspondent to the median  $\hat{\mu}_\sigma$  (*black line*) and CUSUM threshold  $\bar{g}$  (*red dashed line*); Bayesian probability  $\pi_k$  correspondent to the median  $\hat{\mu}_\sigma$  (*black line*), BE threshold (*black dashed line*) and ODP threshold  $F_k(\pi_k, z_k, H_k)$  (*red dash-dotted line*)

chosen HMM) and the solution is achieved recursively, thus facilitating the online implementation.

These results improve over recent formulations of the detection problem for applications in neuroscience [64, 102, 110]. These works, indeed, mainly used spike trains

(i.e., binary sequences) and detected the change points offline by combining data from different states and the knowledge of the entire neuronal activity. In particular, Yu [110] detected changes in the neuronal spiking rate by solving a QD problem, but strict assumptions were made on the class of output observations (which were independent and identically distributed) and the change times  $\bar{T}$  (which followed a geometric distribution).

It is interesting that the solution of the problem (6.3)–(6.6b) results in the threshold-based policy (6.7), which is *adaptive* and *unsupervised*, i.e., the evolution of the threshold is not set *a priori* and depends on the adopted HMM, the distribution functions  $q_0(\cdot)$  and  $q_1(\cdot)$  of the output measurements, and the loss function  $J_0$  [80].

Threshold-based policies have been extensively explored for seizure detection (e.g., [24, 31, 34–36, 42, 45, 61, 73]), but threshold are usually fixed or periodically updated by using heuristic, data-driven paradigms, which might require long training sessions to be more accurate. Also, these policies usually apply to measurements computed out of individual or paired iEEG recording channels, thus requiring the threshold be tuned to the specific location of the electrodes on the brain.

Our detection policy, instead, applies on a Bayesian probability, which is always in the range [0, 1]. Therefore, the dynamics of the *ODP* threshold (6.7) does not explicitly depend on the average amplitude of the measurements in each state and can be applied to different data generated by the same mechanism across multiple trials and conditions, thus improving over the existing heuristic threshold-based policies (e.g., HT, CUSUM, etc.) [80].

It is possible, however, that, although fixed, the threshold in these policies has been chosen “optimally”, i.e., by minimizing a specific cost function. However, the unsupervised optimal approaches proposed thus far (e.g., [5, 32, 33, 37, 38, 54, 68, 91, 93, 95, 99, 104]) usually exploit tools from the theory of machine learning, which means that the optimization process ultimately separates the data in a specific high-dimension feature space, but does not encompass any penalty for performance goals. Consequently, the performances of the resultant detection paradigms follow (and are actually limited by) the formulation of the detection criteria [75].

The proposed framework, instead, defines the required performances *first* by appropriately constructing the loss function to be minimized, and *then* designs the threshold accordingly, thus allowing to trade off between different objects (e.g., low probability of false positives versus low distance between actual and detected change time or low probability of late detection, etc.) depending on the specific application.

Finally, we note that our framework customizes to the specific application and type of observations by exploiting a time-varying, history-dependent HMM, which is estimated offline on training data. For sake of simplicity, we considered here a two-state HMM, with states  $x = 1$  and  $x = 0$  representing the seizure and seizure-free conditions, respectively. However, our approach can be generalized to problems with  $N$  states ( $N > 2$ ) and the correspondent optimal detection policy can be derived as shown in (6.5)–(6.7). This is of particular interest for the seizure onset detection problem, as we recently showed that the brain may transit across several sub-states before and during a seizure event [16, 108] In this case, detecting multiple transitions

or transitions in sub-states preceding the hand-annotated seizure onset would be mostly valuable to early issue warnings or develop seizure-blocking therapies.

It is interesting, though, that a minimal HMM with just two states was enough to detect seizures with very low false positive rates. This is perhaps facilitated by the use of history-dependent generalized linear models (GLMs) to describe the output probabilities  $q_0(\cdot)$  and  $q_1(\cdot)$ . GLMs and maximum likelihood methods have been widely used in the analysis and simulation of neuronal spike trains for several types of neural disorders (e.g., [15, 20, 29, 63, 78, 81, 83, 100, 101]) and provide a flexible framework for both stationary and nonstationary analysis. In our case, the GLM parameters were able to accurately capture changes that occur in the median of the singular values as soon as the seizure starts while requiring a minimal set of training data to be estimated in both conditions.

### 6.8.1 Network-Based Analysis and Singular Value Decomposition

A key aspect of the methodology proposed in [79, 80] is the availability of sequential output measurements whose probability distribution function changes because of a hidden state transition. In the application of our framework to seizure detection we used multi-channel statistics computed out of the iEEG signals as output measurements [75–77]. In particular, the median of the singular values  $\hat{\mu}_\sigma$  of the normalized cross-power based connectivity matrix showed significantly different dynamics in seizure versus non-seizure periods, which indicates that the linear dependencies existing among all the recorded sites of the affected brain and the corresponding brain network topology consistently vary at the transition from interictal to ictal conditions.

The choice of a (linear) network-based output variable is motivated by several drawbacks that have been reported with most of the statistics computed thus far (e.g., [18, 24, 31, 34–36, 40, 41, 49, 52, 58, 59, 61, 69, 73, 93, 99]). These statistics, in fact, are computed from single channels or small subsets of channels from the focal area. However, this requires that the focal areas are known *a priori* with reasonable accuracy, which might be problematic in case of online detection. This requirement, instead, is less stringent when exploiting multi-channel statistics, since it is sufficient that the grid of electrodes is large enough to include the focal areas (which is the case with the current recording schemes) [75].

Also, it has been noted in [48] that nonlinear multi-channel statistics outperform linear single-channel and two-channel measures, but require larger amounts of data and computation, which might be not available during the setup of the detection paradigm.

Finally, it must be noted that all the statistics computed thus far show different patterns in various conditions (e.g., during sleep versus wake state, etc.) and may vary with the specific subject and type of seizure, thus resulting less predictable. These limitations, however, can be addressed in our model-based approach by increasing the number of combined channels and computing simple measures off of large enough

matrices. In particular, more information about the brain network can be derived and both spatial and temporal features can be included in the same model [75].

**Acknowledgments** S. Santaniello was supported by the US National Science Foundation Grant ECCS 1346888. S. V. Sarma was supported by the US National Science Foundation CAREER Award 1055560 and the Burroughs Wellcome Fund CASI Award 1007274. The Bayesian optimal detection framework presented in Sect. 6.2 and Sect. 6.3 was developed in [80] with preliminary results in [79, 82]. The network analysis and the generalized linear model structure presented in Sect. 6.4 and Sect. 6.4.1 were developed in [75] with preliminary results in [77]. Preliminary results with subjects PT-03 and PT-04 were presented in [76].

## References

1. Acharya S, Tenore F, Aggarwal V, Etienne-Cummings R, Schieber MH, Thakor NV (2008) Decoding individuated finger movements using volume-constrained neuronal ensembles in the M1 hand area. *IEEE Trans Neural Syst Rehabil Eng* 16:15–23
2. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
3. Al-Otaibi FA, Hamani C, Lozano AM (2011) Neuromodulation in epilepsy. *Neurosurgery* 69:957–979
4. Alarcon G, Binnie CD, Elwes RD, Polkey CE (1995) Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy. *Electroencephalogr Clin Neurophysiol* 94:326–337
5. Alkan A, Koklukaya E, Subasi A (2005) Automatic seizure detection in EEG using logistic regression and artificial neural network. *J Neurosci Methods* 148:167–176
6. Baier G, Muller M, Stephani U, Muhle M (2007) Characterizing correlation changes of complex pattern transitions: the case of epileptic activity. *Phys Lett A* 363:290–296
7. Baker JJ, Scheme E, Englehart K, Hutchinson DT, Greger B (2010) Continuous detection and decoding of dexterous finger flexions with implantable myoelectric sensors. *IEEE Trans Neural Syst Rehabil Eng* 18:424–432
8. Bansal AK, Truccolo W, Vargas-Irwin CE, Donoghue JP (2012) Decoding 3D reach and grasp from hybrid signals in motor and premotor cortices: spikes, multiunit activity, and local field potentials. *J Neurophysiol* 107:1337–1355
9. Basseville M, Nikiforov IV (1993) *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall, Englewood Cliffs, NJ (USA)
10. Berger JO (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer, New York, NY (USA)
11. Bertsekas DP (2005) *Dynamic Programming and Optimal Control*, vol. I-II, 3rd edn. Athena Scientific, Belmont, MA (USA)
12. de Boer HM, Mula M, Sander JW (2008) The global burden and stigma of epilepsy. *Epilepsy Behav* 12:540–546
13. Bowman AW, Azzalini A (1997) *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford University Press, New York, NY (USA)
14. Brodsky BE, Darkhovsky BS (1993) *Nonparametric Methods in Change-Point Problems*. Kluwer Academic Publishers, Norwell, MA (USA)
15. Brown EN, Barbieri R, Eden UT, Frank LM (2003) Likelihood methods for neural data analysis. In: Feng J (ed) *Computational Neuroscience: A Comprehensive Approach*. CRC, London, UK, pp 253–286
16. Burns SP, Sritharan D, Jouny C, Bergery G, Crone N et al (2012) A network analysis of the dynamics of seizure. 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. San Diego, CA (USA), pp. 4684–4687

17. Cavanna AE, Monaco F (2009) Brain mechanisms of altered conscious states during epileptic seizures. *Nat Rev Neurol* 5:267–276
18. Chavez M, Le Van Quyen M, Navarro V, Baulac M, Martinerie J (2003) Spatio-temporal dynamics prior to neocortical seizures: amplitude versus phase couplings. *IEEE Trans Biomed Eng* 50:571–583
19. Chen X, Hu Y, Fetics BJ, Berger RD, Trayanova NA (2011) Unstable QT interval dynamics precedes ventricular tachycardia onset in patients with acute myocardial infarction: a novel approach to detect instability in QT interval dynamics from clinical ECG. *Circ Arrhythm Electrophysiol* 4:858–866
20. Czanner G, Eden UT, Wirth S, Yanike M, Suzuki WA, Brown EN (2008) Analysis of between-trial and within-trial neural spiking dynamics. *J Neurophysiol* 99:2672–2693
21. Durbin R, Eddy S, Krogh A, Mitchison G (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, UK
22. Elger CE, Schmidt D (2008) *Modern management of epilepsy: a practical approach*. *Epilepsy Behav* 12:501–539
23. Elliott RJ, Aggoun L, Moore JB (1995) *Hidden Markov Models: Estimation and Control*. Springer, New York, NY (USA)
24. Esteller R, Echaz J, D’Alessandro M, Worrell G, Cranstoun S et al (2005) Continuous energy variation during the seizure cycle: towards an on-line accumulated energy. *Clin Neurophysiol* 116:517–526
25. Fisher RS (2012) Therapeutic devices for epilepsy. *Ann Neurol* 71:157–168
26. Fisher RS, van Emde Boas W, Blume W, Elger C, Genton P et al (2005) Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* 46:470–472
27. Fisher RS, Webber WR, Lesser RP, Arroyo S, Uematsu S (1992) High-frequency EEG activity at the start of seizures. *J Clin Neurophysiol* 9:441–448
28. Flint RD, Ethier C, Oby ER, Miller LE, Slutzky MW (2012) Local field potentials allow accurate decoding of muscle activity. *J Neurophysiol* 108:18–24
29. Frank LM, Eden UT, Solo V, Wilson MA, Brown EN (2002) Contrasting patterns of receptive field plasticity in the hippocampus and the entorhinal cortex: an adaptive filtering approach. *J Neurosci* 22:3817–3830
30. Gale JT, Shields DC, Jain FA, Amirmovin R, Eskandar EN (2009) Subthalamic nucleus discharge patterns during movement in the normal monkey and Parkinsonian patient. *Brain Res* 1260:15–23
31. Gardner AB, Worrell GA, Marsh E, Dlugos D, Litt B (2007) Human and automated detection of high-frequency oscillations in clinical intracranial EEG recordings. *Clin Neurophysiol* 118:1134–1143
32. Ghosh-Dastidar S, Adeli H, Dadmehr N (2007) Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. *IEEE Trans Biomed Eng* 54:1545–1551
33. Ghosh-Dastidar S, Adeli H, Dadmehr N (2008) Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans Biomed Eng* 55:512–518
34. Gotman J (1982) Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr Clin Neurophysiol* 54:530–540
35. Gotman J, Gloor P (1976) Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG. *Electroencephalogr Clin Neurophysiol* 41:513–529
36. Grewal S, Gotman J (2005) An automatic warning system for epileptic seizures recorded on intracerebral EEGs. *Clin Neurophysiol* 116:2460–2472
37. Guo L, Rivero D, Dorado J, Rabunal JR, Pazos A (2010) Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J Neurosci Methods* 191:101–109
38. Guo L, Rivero D, Pazos A (2010) Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *J Neurosci Methods* 193:156–163

39. Hu J, Si J, Olson BP, He J (2005) Feature detection in motor cortical spikes by principal component analysis. *IEEE Trans Neural Syst Rehabil Eng* 13:256–262
40. Iasemidis LD, Sackellares JC, Zaveri HP, Williams WJ (1990) Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures. *Brain Topogr* 2:187–201
41. Iasemidis LD, Shiau DS, Sackellares JC, Pardalos PM, Prasad A (2004) Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization techniques. *IEEE Trans Biomed Eng* 51:493–506
42. Jerger KK, Netoff TI, Francis JT, Sauer T, Pecora L et al (2001) Early seizure detection. *J Clin Neurophysiol* 18:259–268
43. Jirsch JD, Urrestarazu E, Le Van P, Olivier A, Dubeau F, Gotman J (2006) High-frequency oscillations during human focal seizures. *Brain* 129:1593–1608
44. Kemere C, Santhanam G, Yu BM, Afshar A, Ryu SI et al (2008) Detecting neural-state transitions using hidden Markov models for motor cortical prostheses. *J Neurophysiol* 100:2441–2452
45. Khan YU, Gotman J (2003) Wavelet based automatic seizure detection in intracerebral electroencephalogram. *Clin Neurophysiol* 114:898–908
46. Kramer MA, Eden UT, Kolaczky ED, Zepeda R, Eskandar EN, Cash SS (2010) Coalescence and fragmentation of cortical networks during focal seizures. *J Neurosci* 30:10076–10085
47. Lee HC, van Drongelen W, McGee AB, Frim DM, Kohrman MH (2007) Comparison of seizure detection algorithms in continuously monitored pediatric patients. *J Clin Neurophysiol* 24:137–146
48. Li D, Zhou W, Drury I, Savit R (2003) Linear and nonlinear measures and seizure anticipation in temporal lobe epilepsy. *J Comput Neurosci* 15:335–345
49. Litt B, Esteller R, Echauz J, D’Alessandro M, Shor R et al (2001) Epileptic seizures may begin hours in advance of clinical onset: a report of five patients. *Neuron* 30:51–64
50. Lombardi F, Porta A, Marzegalli M, Favale S, Santini M et al (2000) Heart rate variability patterns before ventricular tachycardia onset in patients with an implantable cardioverter defibrillator. *Am J Cardiol* 86:959–963
51. Markowitz DA, Wong YT, Gray CM, Pesaran B (2011) Optimizing the decoding of movement goals from local field potentials in macaque cortex. *J Neurosci* 31:18412–18422
52. Martinerie J, Adam C, Le Van Quyen M, Baulac M, Clemenceau S et al (1998) Epileptic seizures can be anticipated by non-linear analysis. *Nat Med* 4:1173–1176
53. McCullagh P, Nelder JA (1990) *Generalized Linear Models*, 2nd edn. CRC, Boca Raton, FL (USA)
54. Meier R, Dittrich H, Schulze-Bonhage A, Aertsen A (2008) Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *J Clin Neurophysiol* 25:119–131
55. Milekovic T, Fischer J, Pistohl T, Ruescher J, Schulze-Bonhage A et al (2012) An online brain-machine interface using decoding of movement direction from the human electrocorticogram. *J Neural Eng* 9:046003
56. Miller P, Katz DB (2010) Stochastic transitions between neural states in taste processing and decision-making. *J Neurosci* 30:2559–2570
57. Minasyan GR, Chatten JB, Chatten MJ, Harner RN (2010) Patient-specific early seizure detection from scalp electroencephalogram. *J Clin Neurophysiol* 27:163–178
58. Mormann F, Kreuz T, Andrzejak RG, David P, Lehnertz K, Elger CE (2003) Epileptic seizures are preceded by a decrease in synchronization. *Epilepsy Res* 53:173–185
59. Mormann F, Lehnertz K, David P, Elger CE (2000) Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients. *Physica D* 144:358–369
60. Niedermeyer E, Lopes da Silva FH (2005) *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th edn. Lippincott Williams & Wilkins, Philadelphia, PA (USA)
61. Osorio I, Frei MG, Wilkinson SB (1998) Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* 39:615–627



62. Page ES (1954) Continuous inspection schemes. *Biometrika* 41:100–115
63. Paninski L (2004) Maximum likelihood estimation of cascade point-process neural encoding models. *Netw* 15:243–262
64. Pillow JW, Ahmadian Y, Paninski L (2011) Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Comput* 23:1–45
65. Ponten SC, Bartolomei F, Stam CJ (2007) Small-world networks and epilepsy: graph theoretical analysis of intracerebrally recorded mesial temporal lobe seizures. *Clin Neurophysiol* 118:918–927
66. Poor HV (1998) Quickest detection with exponential penalty for delay. *Ann Statist* 26:2179–2205
67. Poor HV, Hadjiladias O (2009) *Quickest Detection*. Cambridge University Press, Cambridge, UK
68. van Putten MJ, Kind T, Visser F, Lagerburg V (2005) Detecting temporal lobe seizures from scalp EEG recordings: a comparison of various features. *Clin Neurophysiol* 116:2480–2489
69. Le Van Quyen M, Martinerie J, Navarro V, Baulac M, Varela FJ (2001) Characterizing neurodynamic changes before seizures. *J Clin Neurophysiol* 18:191–208
70. Risinger MW, Engel J Jr, Van Ness PC, Crandall PH (1989) Ictal localization of temporal lobe seizures with scalp/sphenoidal recordings. *Neurology* 39:1288–1293
71. Rummel C, Abela E, Muller M, Hauf M, Scheidegger O et al (2011) Uniform approach to linear and nonlinear interrelation patterns in multivariate time series. *Phys Rev E Stat Nonlin Soft Matter Phys* 83:066215
72. Rummel C, Muller M, Baier G, Amor F, Schindler K (2010) Analyzing spatio-temporal patterns of genuine cross-correlations. *J Neurosci Methods* 191:94–100
73. Saab ME, Gotman J (2005) A system to detect the onset of epileptic seizures in scalp EEG. *Clin Neurophysiol* 116:427–442
74. Sander JW (2003) The epidemiology of epilepsy revisited. *Curr Opin Neurol* 16:165–170
75. Santaniello S, Burns SP, Golby AJ, Singer JM, Anderson WS, Sarma SV (2011) Quickest detection of drug-resistant seizures: an optimal control approach. *Epilepsy Behav* 22:S49–S60
76. Santaniello S, Burns SP, Sarma SV (2012) Automatic seizure onset detection in drug-resistant epilepsy: a Bayesian optimal solution. *IEEE 51st Annual Conference on Decision and Control*. Maui, HI (USA), pp. 3189–3194
77. Santaniello S, Burns SP, Sarma SV (2012) Quickest seizure onset detection in drug-resistant epilepsy. *IEEE American Control Conference*. Montreal, QC (Canada), pp. 4771–4776
78. Santaniello S, Montgomery EB Jr, Gale JT, Sarma SV (2012) Non-stationary discharge patterns in motor cortex under subthalamic nucleus deep brain stimulation. *Front Integr Neurosci* 6:35
79. Santaniello S, Sherman DL, Mirski MA, Thakor NV, Sarma SV (2011) A Bayesian framework for analyzing iEEG data from a rat model of epilepsy. *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Boston, MA (USA), pp. 1435–1438
80. Santaniello S, Sherman DL, Thakor NV, Eskandar EN, Sarma SV (2012) Optimal control-based Bayesian detection of clinical and behavioral state transitions. *IEEE Trans Neural Syst Rehabil Eng* 20:708–719
81. Sarma SV, Eden UT, Cheng ML, Williams ZM, Hu R et al (2010) Using point process models to compare neural spiking activity in the subthalamic nucleus of Parkinson's patients and a healthy primate. *IEEE Trans Biomed Eng* 57:1297–1305
82. Sarma SV, Santaniello S (2011) Quickest detection of state-transition in point processes: application to neuronal activity. *Proceedings of the 18th IFAC World Conference*. Milan, IT, pp. 10021–10026
83. Saxena S, Gale JT, Eskandar EN, Sarma SV (2011) Modulations in the oscillatory activity of the globus pallidus internus neurons during a behavioral task. A point process analysis. *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. Boston, MA (USA), pp. 4179–4182



84. Schalk G, Brunner P, Gerhardt LA, Bischof H, Wolpaw JR (2008) Brain Computer Interfaces (BCIs): detection instead of classification. *J Neurosci Methods* 167:51–62
85. Schiller Y, Cascino GD, Busacker NE, Sharbrough FW (1998) Characterization and comparison of local onset and remote propagated electrographic seizures recorded with intracranial electrodes. *Epilepsia* 39:380–388
86. Schindler K, Amor F, Gast H, Muller M, Stibal A et al (2010) Peri-ictal correlation dynamics of high-frequency (80–200 Hz) intracranial EEG. *Epilepsy Res* 89:72–81
87. Schindler KA, Bialonski S, Horstmann MT, Elger CE, Lehnertz K (2008) Evolving functional network properties and synchronizability during human epileptic seizures. *Chaos* 18:033119
88. Schindler K, Leung H, Elger CE, Lehnertz K (2007) Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG. *Brain* 130:65–77
89. Schmidt D (2009) Drug treatment of epilepsy: options and limitations. *Epilepsy Behav* 15:56–65
90. Shiryaev AN (1963) On optimum methods in quickest detection problems. *Theory Probab Appl* 8:22–46
91. Shoeb A, Edwards H, Connolly J, Bourgeois B et al (2004) Patient-specific seizure onset detection. *Epilepsy Behav* 5:483–498
92. Snyder DL, Miller MI (1991) *Random Point Processes in Time and Space*. Springer, New York, NY (USA)
93. Srinivasan V, Eswaran C, Sriraam N (2007) Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Trans Inf Technol Biomed* 11:288–295
94. Sussillo D, Nuyujukian P, Fan JM, Kao JC, Stavisky SD et al (2012) A recurrent neural network for closed-loop intracortical brain-machine interface decoders. *J Neural Eng* 9:026027
95. Temko A, Thomas E, Marnane W, Lightbody G, Boylan G (2011) EEG-based neonatal seizure detection with Support Vector Machines. *Clin Neurophysiol* 122:464–473
96. Thakor NV, Natarajan A, Tomaselli GF (1994) Multiway sequential hypothesis testing for tachyarrhythmia discrimination. *IEEE Trans Biomed Eng* 41:480–487
97. Thakor NV, Zhu YS, Pan KY (1990) Ventricular tachycardia and fibrillation detection by a sequential hypothesis testing algorithm. *IEEE Trans Biomed Eng* 37:837–843
98. Theodore WH, Spencer SS, Wiebe S, Langfitt JT, Ali A et al (2006) *Epilepsy in North America: a report prepared under the auspices of the global campaign against epilepsy, the International Bureau for Epilepsy, the International League Against Epilepsy, and the World Health Organization*. *Epilepsia* 47:1700–1722
99. Tito M, Cabrerizo M, Ayala M, Jayakar P, Adjouadi M (2009) Seizure detection: an assessment of time- and frequency-based features in a unified two-dimensional decisional space using nonlinear decision functions. *J Clin Neurophysiol* 26:381–391
100. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J Neurophysiol* 93:1074–1089
101. Truccolo W, Hochberg LR, Donoghue JP (2010) Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes. *Nat Neurosci* 13:105–111
102. Ushiba J, Tomita Y, Masakado Y, Komune Y (2002) A cumulative sum test for a peri-stimulus time histogram using the Monte Carlo method. *J Neurosci Methods* 118:207–214
103. Wetherhill GB, Brown DW (1991) *Statistical Process Control: Theory and Practice* Chapman and Hall, London, UK
104. White AM, Williams PA, Ferraro DJ, Clark S, Kadam SD et al (2006) Efficient unsupervised algorithms for the detection of seizures in continuous EEG recordings from rats after brain injury. *J Neurosci Methods* 152:255–266
105. Wilke C, Worrell G, He B (2011) Graph analysis of epileptogenic networks in human partial epilepsy. *Epilepsia* 52:84–93
106. Williams ZM, Neimat JS, Cosgrove GR, Eskandar EN (2005) Timing and direction selectivity of subthalamic and pallidal neurons in patients with Parkinson’s disease. *Exp Brain Res* 162:407–416

107. Worrell GA, Parish L, Cranstoun SD, Jonas R, Baltuch G, Litt B (2004) High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain* 127:1496–1506
108. Yaffe R, Burns S, Gale J, Park HJ, Bulacio J, et al (2012) Brain state evolution during seizure and under anesthesia: a network-based analysis of stereotaxic EEG activity in drug-resistant epilepsy patients. 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. San Diego, CA (USA), pp. 5158–5161
109. Yoon JW, Roberts SJ, Dyson M, Gan JQ (2009) Adaptive classification for Brain Computer Interface systems using sequential Monte Carlo sampling. *Neural Netw* 22:1286–1294
110. Yu AJ (2007) Optimal change-detection and spiking neurons. In: Scholkopf B, Platt JC, Hoffman T (eds) *Advances in Neural Information Processing Systems 19*. MIT Press, Cambridge, MA (USA), pp 1545–1552

**Part II**  
**Biological Network Modelling**

# Chapter 7

## Model Reduction of Genetic-Metabolic Networks via Time Scale Separation

Juan Kuntz, Diego Oyarzún and Guy-Bart Stan

**Abstract** Model reduction techniques often prove indispensable in the analysis of physical and biological phenomena. A successful reduction technique can substantially simplify a model while retaining all of its pertinent features. In metabolic networks, metabolites evolve on much shorter time scales than the catalytic enzymes. In this chapter, we exploit this discrepancy to justify the reduction via time scale separation of a class of models of metabolic networks under genetic regulation. We formalise the concept of a metabolic network and employ Tikhonov's Theorem for singularly perturbed systems. We demonstrate the applicability of our result by using it to address a problem in metabolic engineering: the genetic control of branched metabolic pathways. We conclude by providing guidelines on how to generalise our result to larger classes of networks.

**Keywords** Time scale separation · Model reduction · Genetic-metabolic networks

### 7.1 Introduction

Biological systems often display large discrepancies in the speed at which different processes occur. In such cases, time scale separation is frequently employed to reduce ordinary differential equation (ODE) models of biological phenomena. A classical

---

J. Kuntz · D. Oyarzún · G.-B. Stan (✉)

Department of Bioengineering, Centre for Synthetic Biology and Innovation,  
Imperial College London, London SW7 2AZ, UK  
e-mail: g.stan@imperial.ac.uk

J. Kuntz

e-mail: juan.kuntz08@imperial.ac.uk

D. Oyarzún

Departments of Mathematics, Imperial College London, London SW7 2AZ, UK  
e-mail: d.oyarzun@imperial.ac.uk

example is found in enzyme kinetics [19], whereby the difference between the speed of substrate-enzyme binding and product formation is explicitly used to derive the Michaelis–Menten kinetics.

Another discrepancy is found in genetic-metabolic systems prominent in the field of Metabolic Engineering. These systems describe networks of enzymatic reactions where the concentrations of the catalytic enzymes are dynamically regulated by gene expression. Metabolic reactions occur at rates in the order of seconds or less, while gene expression usually takes between minutes and hours to complete [12]. For this reason, the reduction of models of metabolic networks under genetic control by time scale separation is sometimes used as a stepping stone in the analysis of such models (e.g., [2, 16]). However, the justification behind these reductions is typically limited to qualitative arguments discussing the discrepancy in speed between metabolic and genetic processes. Unfortunately, these arguments sometimes are not sufficient and the reduced model generated does not behave at all like the original (e.g., see [6] for a discussion regarding several models of metabolic networks for which the reduction fails).

In this chapter, we provide sufficient conditions under which reduction via time scale separation of models of metabolic networks under genetic control can confidently be carried out. In Sect. 7.2 we introduce some notation to describe a general class of ODE models of metabolic networks under genetic regulation. In addition, we make certain assumptions on the dynamics of the metabolites. In Sect. 7.3 first we introduce the main ideas behind time scale separation and we consider networks in which the enzyme concentrations are fixed. Then, we present our results regarding the validity of time scale separation as a model reduction tool for metabolic networks. In Sect. 7.4 we conclude the chapter by discussing the plausibility of the assumptions we made throughout the text and the applicability of our results. We illustrate the concepts discussed in the chapter by applying them to the Metabolic Engineering problem presented in Box 7.1.

### **Box 7.1: Genetic control of a branched metabolic network**

The control of metabolic activity of microbes is a long standing problem of the field of Metabolic Engineering. It encompasses the genetic modification of a host organism and its metabolism to optimise or even artificially induce the organism's production of a chemical compound that is of commercial value, e.g., pharmaceuticals, fuels, commodity chemicals, etc., see [25] and references therein. Often, this consists of two steps. First, the selection of a well studied microbial organism as a host (e.g., *E. coli* and *S. cerevisiae*) with some native metabolite that is a precursor to the chemical of interest. Second, the genetic modification of the microbe so that it expresses the enzymes that catalyse the reactions which convert the precursor into the desired molecule [13].

We study a simple instance of the above scenario. Consider the native metabolic pathway in Fig. 7.1a, which converts metabolite 1 into metabolite 3. Suppose that metabolite 2 is a precursor to a chemical of interest, metabolite 4. Suppose that we can design a plasmid that contains the gene coding for enzyme  $e$ , which catalyses the reaction that converts metabolite 2 into metabolite 4 which diffuses across the cell membrane. Since the host requires metabolite 3 to live and grow, we would like to maximise the production of metabolite 4, without greatly disrupting that of metabolite 3. The question now becomes when and how should  $e$  be expressed so that these goals are met.

Consider implementing the controller architecture in Fig. 7.1b. Roughly, if there is an excess of 3, indicating that it is safe to divert resources to the production of 4, then the controller activates the expression of  $e$ , which leads to an increase in the rate of the branch reaction. The branch reaction consumes 2 and, by lowering the concentration of 2, causes a decrease in the production of 3. This drop in production contributes to driving the concentration of 3 back to normal levels. If, on the contrary, the concentration of 3 is initially low, then expression of  $e$  drops and the branch shuts off. In this fashion, 2 is exclusively converted into 3, potentially restoring the concentration of 3 to normal levels.

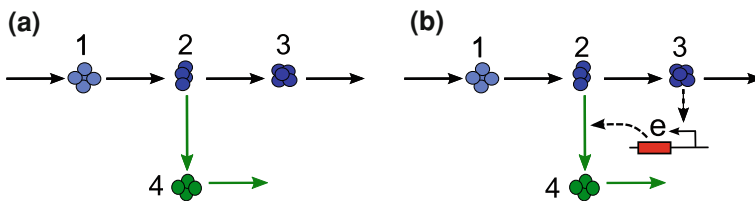
One could describe the above scenario using a model consisting of five ODEs, one of them describing the dynamics of the enzyme concentration and the other four describing the dynamics of the metabolite concentrations. Coarsely, model reductions employing time scale separation consist of grouping model variables into ‘slow’ variables and ‘fast’ variables and then neglecting the dynamics of the fast ones. In our case this grouping would naturally be the four metabolites as the ‘fast’ variables and the enzyme as the ‘slow’ variables. Thus, if applicable, the reduction would permit us to draw conclusions on the behaviour of the network by studying a 1-dimensional model instead of a 5 dimensional model. This would be highly desirable given that the analysis of a 1 dimensional model is straightforward while that of a 5-dimensional model can be exceedingly complicated [8].

## 7.2 Models for Metabolic Reactions Under Genetic Control

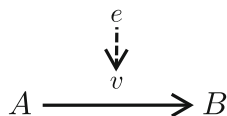
Suppose we have a network of  $n$  metabolites and  $m$  irreversible enzymatic reactions each of which converts a single metabolite into another. Consider the model for the network under genetic regulation

$$\dot{s}(t) = f(s(t), e(t)), \quad s(0) = s_0, \quad (7.1a)$$

$$\dot{e}(t) = g(s(t), e(t)), \quad e(0) = e_0, \quad (7.1b)$$



**Fig. 7.1** Control of a branched metabolic pathway. **a** The native pathway (*black*) converts metabolite 1 into metabolite 3. The synthetic ‘branch’ (*green*) converts the native intermediate, metabolite 2, into a valuable chemical, metabolite 4, and exports it outside the cell. **b** It is possible to implement positive feedback from metabolite 3 to the reaction that converts metabolite 2 into 4 by designing a plasmid coding for the enzyme  $e$ , whose expression is activated by high concentrations of metabolite 3



**Fig. 7.2** An irreversible, enzymatic reaction. The reaction converts metabolite  $A$  into metabolite  $B$  at a rate  $v(s_A, e)$  which depends exclusively on the concentration of the reactant,  $s_A$ , and that of the catalysing enzyme,  $e$

where  $s$  denotes the vector of concentrations of the metabolites and  $e$  denotes the vector of concentrations of the enzymes catalysing the  $m$  reactions in the network. The metabolite dynamics,  $f(\cdot)$ , are defined by the rate at which the reactions consume and produce the different metabolites. The enzyme dynamics,  $g(\cdot)$ , model all the processes involved in enzyme synthesis and degradation.

In this section, we discuss what model (7.1a, 7.1b) represents and make certain assumptions about it. We begin by discussing the kinetics of individual enzymatic reactions. Next, we construct the metabolite dynamics (7.1a) from first principles. We conclude by briefly discussing the enzyme dynamics (7.1b).

### 7.2.1 Enzyme Kinetics

We consider irreversible enzymatic reactions

like the one shown in Fig. 7.2. The reaction converts a single reactant  $A$  into a single product  $B$ . The rate at which the reaction occurs,  $v(s_A, e)$ , depends exclusively on the concentration of the reactant,  $s_A$ , and the concentration of the catalysing enzyme,  $e$ .

**Assumption 1** The reaction rate,  $v(s_A, e)$ , is smooth and globally Lipschitz continuous. For any given constant enzyme concentration  $e > 0$ , we assume that  $v(\cdot, e)$  is bounded, that

$$\frac{\partial v(s_A, e)}{\partial s_A} > 0, \quad \forall s_A \neq 0,$$

and that  $v(\cdot, e)$  is *positive definite*, that is,

$$v(0, e) = 0, \quad v(s_A, e) > 0, \quad \forall s_i \neq 0.$$

We denote its least upper bound with

$$\lim_{s_A \rightarrow +\infty} v(s_A, e) = \hat{v}(e).$$

At a network level we need to distinguish between different reactions. To do this, we write  $v_{A \rightarrow B}$  and  $e_{A \rightarrow B}$  to refer to the rate and the concentration of the catalysing enzyme of the reaction with reactant  $A$  and product  $B$ .

Our assumptions on the kinetics are satisfied by a wide range enzyme kinetics proposed in the literature [3] (e.g., Michaelis–Menten and Hill type kinetics). Essentially, they state that:

- (Positive definite) If there are no reactant molecules present, the reaction rate is zero. If there are some reactant and some enzyme molecules present, the reaction rate is non-zero.
- (Strictly increasing) If there are some enzyme molecules present, then the more reactant molecules present, the faster the reaction rate.
- (Bounded) Enzymes have a limited number of active sites that reactants can bind to. Thus, given a fixed number of enzyme molecules, the reaction rate cannot exceed the maximum rate achieved when all the enzymes' active sites are bound by the reactants.

Implicit in our definition of the reaction rates is the assumption that they are time invariant. It is well known that the rate of a reaction depends on the temperature and pressure of the medium in which the reaction is taking place. Hence, assuming time invariance of the reaction kinetics is equivalent to assuming that the cytoplasm can be approximated to be isobaric and isothermal. This is a common assumption in the literature on ODE models of biochemical reactions [3, 7].

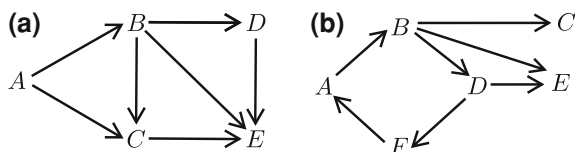
### 7.2.2 Metabolic Model

Assuming that the cytoplasm may be approximated to be an isovolumetric and spatially homogeneous medium [7], the law of mass balance applied to the concentration of metabolite number  $i$  yields

$$\dot{s}_i(t) = P_i(t) - C_i(t) + I_i(t) - E_i(t), \quad (7.2)$$



**Fig. 7.3** Acyclicity in networks. **a** The network is acyclic. **b** The network is not acyclic;  $A, B, D, F$  form a cycle



where  $P_i$  denotes the rate at which  $s_i$  is produced by the considered genetic-metabolic network,  $C_i$  the rate at which  $s_i$  is consumed by the network,  $I_i$  the rate at which  $s_i$  enters the network from outside and  $E_i$  the rate at which  $s_i$  leaves the network. From now on, we use the convention  $v_{i \rightarrow j} \equiv 0$  if there is no reaction that converts metabolite  $i$  into  $j$ .

A metabolite is produced (consumed) by the reactions of which it is the product (reactant). We limit our attention to networks whose metabolites can be ordered in such a way that the following condition is satisfied.

**Condition 1** For any  $i$ , if  $j > i$  then  $v_{j \rightarrow i} \equiv 0$ . In other words, metabolite  $i$  is not the product of any reaction whose reactant is metabolite  $i + 1, i + 2, \dots, n$ .

Condition 1 has a simple graphical interpretation. Consider the directed graph whose vertices represent the metabolites and whose edges represent the transfer of mass (via reactions) from one metabolite to another. Condition 1 is equivalent to the graph being **acyclic**, that is, starting at any given vertex one cannot return to that same vertex by following the edges, see Fig. 7.3. Examples of such networks can be found in the amino acid biosynthesis pathways of *E. coli* [24].

Let  $N_{i,i \rightarrow j}$  denote the *stoichiometric coefficient* of  $i$  in reaction  $i \rightarrow j$ , that is, the number of molecules of  $i$  involved in reaction  $i \rightarrow j$ . If Condition 1 holds, we can write the rates of production as

$$P_1(t) := 0, \quad P_i(t) := \sum_{j=1}^{i-1} N_{i,j \rightarrow i} v_{j \rightarrow i}(s_j, e_{j \rightarrow i}), \quad i = 2, 3, \dots, n, \quad (7.3)$$

and the rates of consumption as

$$C_n(t) := 0, \quad C_i(t) := \sum_{j=i+1}^n N_{i,i \rightarrow j} v_{i \rightarrow j}(s_i, e_{i \rightarrow j}), \quad i = 1, 2, \dots, n-1. \quad (7.4)$$

**Assumption 2** The import rates are constant,  $I_i(t) := I_i \geq 0 \forall i$ . The export rate of a metabolite  $i$ , if it exists, is a smooth, globally Lipschitz continuous, positive definite, bounded function of its concentration such that

$$\frac{\partial E_i(s_i)}{\partial s_i} > 0, \quad \forall s_i \neq 0.$$

We denote its least upper bound with

$$\lim_{s_i \rightarrow +\infty} E_i(s_i) = \hat{E}_i \quad \forall s_i \neq 0.$$

One can use the import and export rates to model a variety of phenomena. For instance, they may represent the rates at which the metabolites flow in and out of the cell. Or the rates at which the metabolites are consumed/produced by other metabolic pathways inside the cell. Additionally, one may use the export rates to circumvent the isovolumetric assumption and model dilution. In any of these cases the physical interpretations of Assumption 2 are similar to those we made regarding the assumptions on the enzyme kinetics (Assumption 1). In addition, assumptions of the type of Assumption 2 are common in the systems biology literature (for example, see [4, 17]) and for this reason we shall not discuss them any further.

We can now rewrite the metabolite dynamics,  $f(s, e)$ , in the model (7.1a, 7.1b) as

$$\begin{aligned} \dot{s}_1 &= I_1 - E_1(s_1) - \sum_{j=2}^n N_{1,1 \rightarrow j} v_{1j}(s_1, e_{1 \rightarrow j}), \\ \dot{s}_i &= I_i + \sum_{j=1}^{i-1} N_{i,j \rightarrow i} v_{j \rightarrow i}(s_j, e_{j \rightarrow i}) \\ &\quad - E_i(s_i) - \sum_{j=i+1}^n N_{i,i \rightarrow j} v_{i \rightarrow j}(s_i, e_{i \rightarrow j}), \quad i = 2, 3, \dots, n-1, \\ \dot{s}_n &= I_n + \sum_{j=1}^{n-1} N_{n,j \rightarrow n} v_{j \rightarrow n}(s_j, e_{j \rightarrow n}) - E_n(s_n). \end{aligned} \quad (7.5)$$

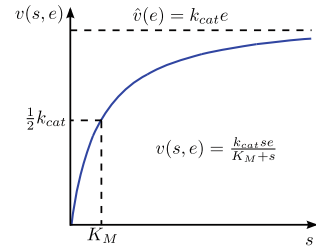
### Box 7.2: Metabolic model

In our example network we assume that all reactions follow Michaelis-Menten kinetics

$$v_{1 \rightarrow 2} := \frac{k_{cat1}s_1}{K_{M1} + s_1} e_{1 \rightarrow 2}, \quad v_{2 \rightarrow 3} := \frac{k_{cat2}s_2}{K_{M2} + s_2} e_{2 \rightarrow 3}, \quad v_{2 \rightarrow 4} := \frac{k_{cat3}s_2}{K_{M3} + s_2} e_{2 \rightarrow 4}.$$

It is straightforward to verify that Michaelis–Menten kinetics satisfy Assumption 1, see Fig. 7.4.

**Fig. 7.4** Michaelis Menten kinetics. The kinetics are strictly increasing, positive definite and bounded



The network in Fig. 7.1 has a single import rate,  $I_1$  and two export rates,  $E_3$  and  $E_4$ . We assume that the export rates may also be described by Michaelis-Menten functions

$$E_3 := \frac{\hat{E}_3 s_3}{K_{O3} + s_3}, \quad E_4 := \frac{\hat{E}_4 s_4}{K_{O4} + s_4}.$$

Hence, we obtain the metabolite dynamics

$$\dot{s}_1 = I_1 - \frac{k_{cat1}s_1}{K_{M1} + s_1}e_{1 \rightarrow 2}, \quad (7.6a)$$

$$\dot{s}_2 = \frac{k_{cat1}s_1}{K_{M1} + s_1}e_{1 \rightarrow 2} - \frac{k_{cat2}s_2}{K_{M2} + s_2}e_{2 \rightarrow 3} - \frac{k_{cat3}s_2}{K_{M3} + s_2}e_{2 \rightarrow 4}, \quad (7.6b)$$

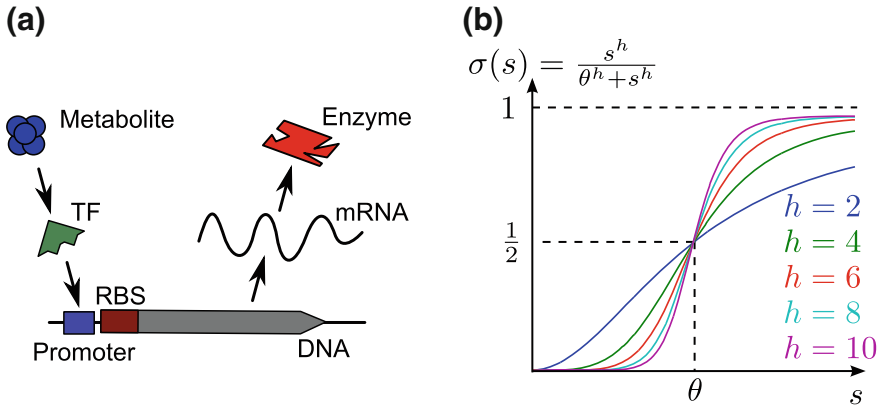
$$\dot{s}_3 = \frac{k_{cat2}s_2}{K_{M2} + s_2}e_{2 \rightarrow 3} - \frac{\hat{E}_3 s_3}{K_{O3} + s_3}, \quad (7.6c)$$

$$\dot{s}_4 = \frac{k_{cat3}s_2}{K_{M3} + s_2}e_{2 \rightarrow 4} - \frac{\hat{E}_4 s_4}{K_{O4} + s_4}. \quad (7.6d)$$

From Fig. 7.1a it is easy to see that our network satisfies Condition 1, i.e., it is acyclic. To simplify future computations, we choose  $k_{cati} = k_{cat} = 32 \text{ s}^{-1}$ ,  $K_{Mi} = K_M = 4.7 \mu\text{Ms}^{-1} \forall i$  and  $e_{1 \rightarrow 2} = e_{2 \rightarrow 3} = e_N = 200 \text{ nM}$ . These values are representative of reactions in the tryptophan pathway (extracted from the BRENDA database [18], EC number 5.3.1.24). We also assume that  $\hat{E}_3 = \hat{E}_4 = k_{cat}e_N$ ,  $K_{O3} = K_{O4} = K_M$  and use the shorthand  $e := e_{2 \rightarrow 4}$ .

### 7.2.3 Enzymatic Model

The enzyme dynamics,  $g(\cdot)$ , are a lumped representation of all the processes involved in enzyme synthesis and removal. Synthesis encompasses the transcription of genes encoding the enzymes by RNA polymerases into mRNA strands and the



**Fig. 7.5** The implementation of feedback via promoter design. **a** Metabolite 3 induces a conformation change on the transcription factor, which then binds to the promoter of the gene coding for  $e$  and activates its expression. **b** Hill functions with different Hill coefficients, note that their range is  $[0, 1)$ , hence  $k_0 + k_1$  represent the maximum rate of expression of  $e$

translation of these by ribosomes into proteins that later fold into the actual enzyme. Most enzyme–enzyme and metabolite–enzyme interactions occur in synthesis, specifically in transcription. In particular, metabolites often act as, or activate/deactivate transcription factors that inhibit or activate the transcription of genes coding for other enzymes. Removal typically, includes enzyme degradation by the cell and dilution due to cell growth.

To keep this exposition general, we shall not define the function  $g(\cdot)$  explicitly. We will only make the following minimal assumptions.

**Assumption 3** The enzymes dynamics  $g(\cdot)$  are smooth and globally Lipschitz continuous.

Enzyme degradation and dilution are typically modelled as linear functions of the enzyme concentration [1]. Synthesis is usually modelled as the sum of a constant (or basal) expression rate a set of sigmoids (e.g., Hill functions) representing the activating or repressing effects of the transcription factors on the enzyme expression [2, 14, 15]. These are all smooth and globally Lipschitz continuous functions. The enzyme dynamics,  $g(\cdot)$ , are a linear combination of these and, thus, are also a smooth and globally Lipschitz continuous. For this reason, Assumption 3 holds for a significant portion of the models presented in the literature.

**Box 7.3: Enzymatic model**

Consider the controller for the branched metabolic pathway discussed in Box 7.1. Implementing such a controller can be achieved, for example, by designing

the promoter of the gene coding for  $e$  such that 3 binds to some transcription factor that activates the transcription of  $e$ , see Fig. 7.5a. We model the expression of the branch enzyme  $e$  as

$$\dot{e} = k_0 + k_1\sigma(s_3) - \gamma e, \quad \sigma(x) := \frac{x^h}{\theta^h + x^h}. \quad (7.7)$$

This model comes from the balance between protein synthesis and degradation. We consider a first order removal process with kinetic constant  $\gamma$ , which accounts for the aggregate effect of degradation and dilution by cell growth [1]. The synthesis term,  $k_0 + k_1\sigma(s_3)$ , describes both transcription and translation of  $e$ . The parameter  $k_0$  represents the leaky expression of the enzyme that occurs regardless whether the gene is activated or repressed, while  $k_1$  represents the compound effect of transcription and translation when the gene is fully expressed. The function  $\sigma(\cdot)$  takes values in  $[0, 1)$  and depends on the specific molecular mechanisms underlying the interactions both between metabolite 3 and the transcription factor and those between the transcription factor and the promoter of the gene coding for the enzyme. Typically, these types of interactions are modelled as sigmoidal (or Hill) functions [2, 15], see Fig. 7.5b.

Following [15], we chose the parameter values  $k_0 = 0.03$  nM,  $k_1 = 100k_0$ ,  $\gamma = 2 \times 10^{-4}$  s<sup>-1</sup>,  $\theta = 0.2$   $\mu$ M and  $h = 2$ .

## 7.3 Model Reduction via Time Scale Separation

In this section we present our results regarding time scale separation in genetic-metabolic systems. We first consider the behaviour of metabolic networks when the enzyme concentrations are kept fixed in time. There are two reasons behind this. First, it is a prerequisite of the time scale separation results regarding networks with varying enzyme concentrations. Second, the study itself is instructive with regard to understanding the behaviour of the networks. After this, we introduce abstractly the main ideas of time scale separation and give our results justifying the time scale separation based reduction of the networks.

### 7.3.1 Metabolic Networks with Constant Enzyme Concentrations

Suppose that the enzyme concentrations are positive constants, i.e.,  $e(t) \equiv e \in \mathbb{R}_{>0}^m$ . We find it convenient to rewrite (7.5) as

$$\dot{s}_i = g_i(s_1, \dots, s_{i-1}, e) - h_i(s_i, e), \quad i = 1, 2, \dots, n, \quad (7.8)$$

where  $g_1 := I_1$ ,

$$g_i(s_1, \dots, s_{i-1}, e) := I_i + \sum_{j=1}^{i-1} N_{i,j \rightarrow i} v_{j \rightarrow i}(s_j, e_{j \rightarrow i}), \quad i = 2, 3, \dots, n,$$

and  $h_n(s_n) := E_n(s_n)$ ,

$$h_i(s_i, e) := E_i(s_i) + \sum_{j=i+1}^n N_{i,i \rightarrow j} v_{i \rightarrow j}(s_i, e_{i \rightarrow j}), \quad i = 1, 2, \dots, n-1.$$

The function  $g_i(\cdot) \geq 0$  represents the total rate of increase (via both import and production) of the concentration of metabolite  $i$ . Similarly, the function  $h_i(\cdot) \geq 0$  represents the total rate of decrease (via both export and consumption) of the concentration of metabolite  $i$ . To avoid pathological scenarios in which the concentration of a given metabolite grows unbounded because there is no process that removes the metabolite, we impose the following condition.

**Condition 2** *Every metabolite has at least one reaction that consumes it, or it has a non-trivial export term. In other words, for all  $i$ ,  $E_i \neq 0$  or there exists a  $j$  such that  $v_{i \rightarrow j} \neq 0$ . Thus,  $h_i \neq 0$  for all  $i$ .*

The non-zero  $E_i$  and  $v_{i \rightarrow j}$  functions (if they exist, and Condition 2 ensures at least one does exist for all  $i = 1, 2, \dots, n$ ) are bounded, strictly increasing, positive definite functions of  $s_i$ . Hence,  $h_i$ , which is a sum of these functions, is also positive definite and strictly increasing with  $s_i$  and it maps from  $\mathbb{R}_{\geq 0}$  to  $[0, \hat{h}_i(e))$ , where

$$\hat{h}_i(e) := \hat{E}_i + \sum_{j=i+1}^n N_{i,i \rightarrow j} \hat{v}_{i \rightarrow j}(e_{i \rightarrow j}).$$

We now examine the conditions under which system (7.8) an equilibrium  $\bar{s}$ . By definition,  $g_1 \equiv I_1 \geq 0$ , thus  $\dot{s}_1 = 0$  implies

$$h_1(\bar{s}_1, e) = I_1.$$

This equation has a solution if and only if the  $I_1$  is in the range of the function  $h_1(\bar{s}_1, e)$ . In other words, we require  $\hat{h}_1(e) > I_1$ . In addition,  $h_1(\bar{s}_1, e)$  is a strictly increasing function of  $\bar{s}_1$ , thus if a solution exists it is unique. Now, if we assume that  $\bar{s}_1, \dots, \bar{s}_{i-1}$  exist, then  $\dot{s}_i = 0$  implies

$$h_i(\bar{s}_i, e) = g_i(\bar{s}_1, \dots, \bar{s}_{i-1}, e).$$

Similarly as before, the equation has a solution if and only if the constant  $g_i(\bar{s}_1, \dots, \bar{s}_{i-1}, e)$  is in the range of the function  $h_i(\bar{s}_i, e)$ , i.e., if  $\hat{h}_i(e) > g_i(\bar{s}_1, \bar{s}_2, \dots, \bar{s}_{i-1}, e)$ . In addition,  $h_i(\bar{s}_i, e)$  is a strictly increasing function of  $\bar{s}_i$ . Hence if a solution exists it is unique.

Thus, by induction, an equilibrium exists if and only if the following condition is satisfied.

**Condition 3** *The vector of constant enzymes  $e$  is such that  $\hat{h}_i(e) > g_i(\bar{s}_1, \dots, \bar{s}_{i-1}, e) \forall i = 1, 2, \dots, n$ .*

Furthermore, by monotonicity of the  $h_i$ s, if the equilibrium exists it is unique.

Condition 3 is important and has an intuitive interpretation. Regard the metabolites in the network as large water tanks, their concentrations as the water level in the tanks, the reactions as pipes connecting the tanks and the reaction rates as the rate of flow of water through the pipes. In this context, the enzymes may be regarded as valves whose concentrations modulate the resistance to flow through them. Then  $g_i(\cdot)$  may be interpreted as the rate at which water enters the  $i$ th tank through the incoming pipes and  $h_i(\cdot)$  as the rate at which it leaves through the outgoing pipes. The monotonicity of  $h_i$  can be interpreted as ‘the more volume of water in the tank, the greater the water pressure and thus the bigger the rate at which the water is pushed out of the tank through the outgoing pipes’. Condition 3 simply ensures that the outgoing pipes are ‘sufficiently large’ in the sense that the maximum rate at which water can escape the tank is higher than the equilibrium rate at which water enters.

Condition 1, that the network is acyclic, implies that there is no chain of reactions that convert metabolite  $i$  into metabolites  $1, 2, \dots, i - 1$ . Thus, if metabolites  $1, 2, \dots, i - 1$  are at their equilibrium concentrations, they will remain there forever irrespective of what is happening to the concentrations of metabolites  $i, i + 1, \dots, n$ . So, if Condition 3 does not hold for a given metabolite  $i$  and metabolites  $1, 2, \dots, i - 1$  are at their equilibrium concentrations, then metabolite  $i$  will simply accumulate and its concentration will grow unbounded.

#### Box 7.4: Network fluxes

Consider Condition 3 applied to the network in Fig. 7.1a

$$\begin{aligned} \hat{v}_{1 \rightarrow 2}(e_N) &> I_1, & \hat{v}_{2 \rightarrow 3}(e_N) + \hat{v}_{2 \rightarrow 4}(e) &> v_{1 \rightarrow 2}(\bar{s}_1, e_N), \\ \hat{E}_3 &> v_{2 \rightarrow 3}(\bar{s}_2, e_N), & \hat{E}_4 &> v_{2 \rightarrow 4}(\bar{s}_2, e). \end{aligned}$$

By definition, all the reaction rates are non-negative, so  $\hat{v}_{2 \rightarrow 3}(e_N) + \hat{v}_{2 \rightarrow 4}(e) \geq \hat{v}_{2 \rightarrow 3}(e_N)$ . Also note that because  $\bar{s}$  is an equilibrium

$$I_1 = v_{1 \rightarrow 2}(\bar{s}_1, e_N) = v_{2 \rightarrow 3}(\bar{s}_2, e_N) + v_{2 \rightarrow 4}(\bar{s}_2, e).$$

In Box 7.2 we assumed that

$$\hat{v}_{1 \rightarrow 2}(e_N) = \hat{v}_{2 \rightarrow 3}(e_N) = \hat{E}_3 = \hat{E}_4 = k_{cat} e_N.$$

So, Condition 3 is satisfied for any positive enzyme concentration, that is  $e \in (0, +\infty)$ , if and only if  $k_{cat} e_N > I_1$ .

It can be shown that the fulfilment of Condition 3 does not just imply that the network has a unique equilibrium, it also implies that the equilibrium is stable.

**Lemma 1** *Assume that the metabolic network is such that Conditions 1 and 2 are satisfied and Assumptions 1 and 2 hold. If the enzyme concentrations are fixed in time at some value such that Condition 3 is satisfied, then (7.8) has a unique equilibrium which is globally asymptotically stable.*

The proof of the above lemma can be found in Appendix 2.

### 7.3.2 Time Scale Separation

Time scale separation is applicable to systems that can be written as

$$\varepsilon \dot{z} = f(x, z), \quad z(0) = z_0 \tag{7.9a}$$

$$\dot{x} = g(x, z), \quad x(0) = x_0 \tag{7.9b}$$

where the components of  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ ,  $g: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$  are in the same order of magnitude for all  $(x, z) \in \mathbb{R}^{n+m}$  and  $0 < \varepsilon \ll 1$  is a small positive real number. The characterising feature of these systems is that the dynamics of some of the state variables ( $z$ ) are multiple orders of magnitude faster than those of the other state variables ( $x$ ), i.e.,  $\dot{z} = f(x, z)/\varepsilon \gg g(x, z) = \dot{x}$ . Suppose that during a small interval of time within which the value of the slow variables ( $x$ ) remain approximately constant, the fast variables ( $z$ ), which are evolving hundreds/thousands times faster, reach some steady state or *quasi-steady state*. If we assume that the dynamics of the variables  $z$  reach this steady-state very quickly (almost instantaneously at the time scale of the slow variables  $x$ ), then we can assume that, at the time scale of the slow variables  $x$ ,  $\dot{z} = 0$  or, equivalently, that

$$f(x, z) = 0.$$

Suppose that the above has a unique root  $z = \phi(x)$ , i.e.,  $f(x, \phi(x)) = 0$  for all  $x$ . Then, at the time scale of the slow variables  $x$ , one can focus on studying the *reduced* dynamical system

$$\dot{\bar{x}} = g(\bar{x}, \phi(\bar{x})), \quad \bar{x}(0) = x_0 \tag{7.10a}$$

$$\bar{z} = \phi(\bar{x}), \tag{7.10b}$$



instead of the original system (7.9).

Notice that in contrast with the fast variable  $z$  of the original system (7.9), which starts at time 0 from a given  $z_0$ , the fast variable  $\bar{z}$  of the reduced system (7.10) is not free to start from  $z_0$  and there may be a large discrepancy between its initial value,  $\phi(x_0)$ , and  $z_0$ . Thus, there must at least be a short period of time where the behaviour of reduced system does not approximate well that of the complete system.

Before carrying out the above reduction, we need to address a number of outstanding issues. For instance, does a quasi-steady state exist? Is it unique? If not, which quasi-steady state should be used in the reduction? Do the fast variables of the complete system always tend to their quasi-steady state?

Theorem 2 (in the appendix), known as *Tikhonov's Theorem*, partly answers these questions by providing sufficiency conditions under which the behaviour of the original system (7.9) is well approximated by that of the reduced system (7.10). More specifically, if its assumptions are satisfied, Tikhonov's Theorem ensures that after some period of time of order  $\varepsilon \ln(1/\varepsilon)$ , during which the initial discrepancy between  $z$  and  $\bar{z}$  dies out, the norm difference between the trajectory of the complete system (7.9) and that of the reduced system (7.10) remains of order  $\varepsilon$  and no more.

### 7.3.3 Sufficiency Conditions for Time Scale Separation

To be able to state our results regarding time scale separation in genetic-metabolic systems, we must first re-write the network model (7.1a, 7.1b) in the same form as (7.9). Usually, this involves some, possibly complicated, change of variables. However, in the case of genetic-metabolic networks this is not necessary; the 'fast' variables are the metabolite concentrations while the 'slow' variables are the enzyme concentrations. Thus all that must be done is to scale the variables so that the new metabolite dynamics,  $f(\cdot)$ , and the enzyme dynamics,  $g(\cdot)$ , are of the same order of magnitude and all the normalising constants are grouped into a parameter  $\varepsilon$  multiplying  $\hat{s}$ . A systematic way to do this is to *non-dimensionalise* the network model (7.1a, 7.1b), which consists of performing a set of variable substitutions such that the new variables have no physical dimensions associated with them [11].

#### Box 7.5: Non-dimensionalisation

Consider substituting the variables of our network model (Eqs. (7.6) in Box 7.2 and (7.7) in Box 7.3) with

$$z := \frac{s}{K_M}, \quad x := \frac{e}{\hat{e}}, \quad \tau := \gamma t, \quad \hat{e} := \frac{k_0 + k_1}{\gamma}. \quad (7.11)$$

Notice that the new variables  $(x, z)$  have no physical units associated with them. After re-arranging we get

$$\varepsilon \frac{dz_1}{d\tau} = \tilde{I} - \frac{z_1}{1+z_1} \quad (7.12a)$$

$$\varepsilon \frac{dz_2}{d\tau} = \frac{z_1}{1+z_1} - \frac{z_2}{1+z_2} - \frac{\hat{e}}{e_N} \frac{z_2 x}{1+z_2} \quad (7.12b)$$

$$\varepsilon \frac{dz_3}{d\tau} = \frac{z_2}{1+z_2} - \frac{z_3}{1+z_3} \quad (7.12c)$$

$$\varepsilon \frac{dz_4}{d\tau} = \frac{\hat{e}}{e_N} \frac{z_2 x}{1+z_2} - \frac{z_4}{1+z_4} \quad (7.12d)$$

$$\frac{dx}{d\tau} = \frac{k_0}{k_0+k_1} + \frac{k_1}{k_0+k_1} \sigma^*(z_3) - x \quad (7.12e)$$

where  $\tilde{I} = \frac{I_1}{k_{cat}e_N}$ ,  $\sigma^*(z_3) := \sigma(K_M z_3)$  and  $\varepsilon = \frac{KM\gamma}{k_{cat}e_N} \approx 1.5 \times 10^{-4}$ .

We can now state our results regarding time scale separation in metabolic networks under genetic regulation. The proofs for the following lemma and theorem may be found in Appendix 2.

**Lemma 2** *Suppose that (7.1a, 7.1b) is such that Conditions 1, 2 and Assumptions 1–3 hold. Consider a non-dimensionalised version of (7.1a, 7.1b)*

$$\varepsilon \dot{s}(t) = f(s(t), e(t)), \quad s(0) = s_0 \quad (7.13a)$$

$$\dot{e}(t) = g(s(t), e(t)), \quad e(0) = e_0 \quad (7.13b)$$

Then, the unique solution of (7.13),  $[s(t) \ e(t)]^T$ , exists for all  $t \geq 0$ . In addition, let  $A$  denote the subset of  $\mathbb{R}_{>0}^m$  whose elements are such that Condition 3 holds. There exists a unique function  $\phi: A \rightarrow \mathbb{R}^n$  such that  $f(\phi(e), e) = 0$  for all  $e \in A$ . In addition,  $\phi(\cdot)$  is continuously differentiable. Consider the reduced system

$$\dot{\bar{e}}(t) = g(\phi(\bar{e}(t)), \bar{e}(t)), \quad \bar{e}(0) = e_0. \quad (7.14)$$

Suppose that there exists a compact set  $B \subseteq A$  that is forward invariant with respect to (7.14). Then, if  $e_0 \in B$ , (7.14) has a unique solution  $\bar{e}(t) \in B$  for all  $t \geq 0$ .

**Theorem 1** *Suppose that the assumptions of Lemma 2 are satisfied and that  $e_0 \in B$ . Then, for any finite time  $T \geq 0$*

$$e(t) = \bar{e}(t) + O(\varepsilon) \quad (7.15)$$

holds for all  $t \in [0, T]$  and there exists a time  $t_1 \geq 0$ ,  $O(\varepsilon \ln(1/\varepsilon))$ , such that

$$s(t) = \bar{s}(t) + O(\varepsilon), \quad (7.16)$$

where  $\bar{s}(t) := \phi(\bar{e}(t))$ , holds for all  $t \in [t_1, T]$ .

### Box 7.6: Model Reduction

As discussed in Box 7.4, Condition 3 is satisfied for all values of  $e \in (0, +\infty)$ , or equivalently  $x \in (0, +\infty)$ , if and only if  $\tilde{I} < 1$ . Suppose that this is so and define  $A := (0, +\infty)$ . Then, for any  $x \in A$ , the non-dimensionalised model (7.12) has the unique root

$$\phi_1(x) = \frac{\tilde{I}}{1 - \tilde{I}}, \quad \phi_2(x) = \phi_3(x) = \frac{\tilde{I}}{\frac{\hat{e}}{e_N}x + 1 - \tilde{I}}, \quad \phi_4(x) = \frac{\tilde{I}}{\frac{e_N}{\hat{e}}\frac{1}{x} + 1 - \tilde{I}}.$$

Thus, the reduced model is given by

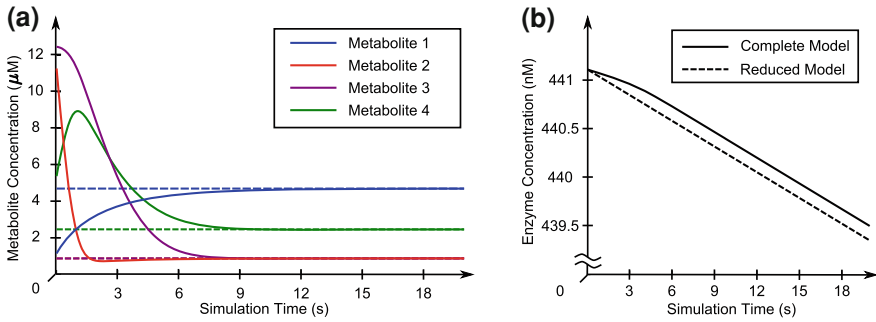
$$\dot{\bar{x}} = \frac{k_0}{k_0 + k_1} + \frac{k_1}{k_0 + k_1} \sigma^*(\phi_3(\bar{x})) - \bar{x}, \quad \dot{\bar{z}} = \phi(\bar{x}). \quad (7.17)$$

To satisfy the premise of Theorem 1, and thus justify the reduction, all that remains to be done is to find a compact subset of  $A$  that is forward invariant with respect to (7.17). Given that  $\sigma(x)^* \in [0, 1)$  for all  $x \in [0, +\infty)$  we have that

$$\frac{k_0}{k_0 + k_1} - \bar{x} \leq \dot{\bar{x}} \leq 1 - \bar{x}. \quad (7.18)$$

From the above it is straightforward to see that  $[\frac{k_0}{k_0+k_1}, 1]$  is a compact subset of  $A$  that is forward invariant with respect to (7.17). Suppose that  $x_0 \in [\frac{k_0}{k_0+k_1}, 1]$ , or, equivalently,  $e_0 \in [\frac{k_0}{\gamma}, \frac{k_0+k_1}{\gamma}]$ . Then, using the substitutions in (7.11), Theorem 1 implies that the norm of the difference between the enzyme trajectory of the our original model (7.6), (7.7) and that of the reduced model (7.17) will be of order 0.037 nM and that, after a short period of time (of order 1.3 ms), the norm of the difference between metabolite trajectory of both models will be of order 0.69 nM, see Fig. 7.6.

The main benefit of carrying out this reduction, is that it can often be considerably easier to extract analytical results from the lower dimensional reduced model than from the higher dimensional original model. This is particularly obvious in our example given that in Box 7.6 we reduced a 5-dimensional model to a 1-dimensional model.

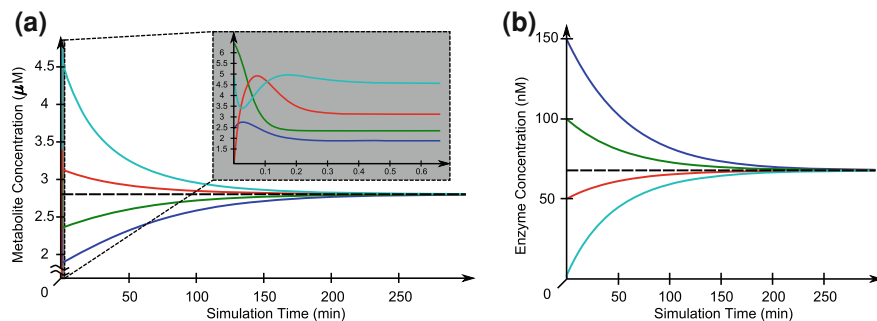


**Fig. 7.6** Model reduction. The plots were generated using MATLAB and show the first few seconds of a simulation of a single trajectory of both the original and reduced models, (7.17) and (7.12), respectively. These were generated using  $I_1 = \frac{1}{2}k_{cat}e_N$  (thus  $\tilde{I}_1 = \frac{I}{k_{cat}e_N} = 1/2 < 1$ ). **a** The trajectory of the metabolites of the complete model (solid lines) converges rapidly to that of the reduced model (dashed lines). **b** The trajectory of the enzyme of the complete model remains a fraction of a nano molar away from that of the reduced model

### Box 7.7: Global stability of the reduced model

The dynamics of the reduced system (7.17),  $g(\bar{x}, \phi(\bar{x}))$ , is a strictly decreasing function of  $\bar{x}$ . This follows from the fact that  $\phi_3$  is a decreasing function of its argument while  $\sigma^*$  is an increasing function of its argument. So  $\sigma^*(\phi_3(x))$  is a decreasing function of  $x$ . In addition, due to inequality (7.18),  $g(0, \phi(0)) \geq \frac{k_0}{k_0+k_1} > 0$  and  $g(1, \phi(1)) \leq 0$ . This, together with the fact that  $g(\phi(x), x)$  is a continuous function of  $x$  implies that the reduced model has a unique equilibrium  $\bar{x}_{eq} \in [0, 1]$ . Lastly, the reduced model is a 1 dimensional system, hence, the fact that  $g(\bar{x}, \phi(\bar{x}))$  is strictly decreasing in  $\bar{x}$ , implies that the unique equilibrium is globally asymptotically stable, see Fig. 7.7.

In conclusion, such a controller architecture ensures that the network has a unique steady state to which the concentrations of the metabolites and of the enzyme always tend to, regardless of initial their values. In addition, by modifying the promoter parameters (in particular, the basal expression  $k_0$  and promoter strength  $k_1$ ) one can move the steady state to a more desirable location (e.g., maximise the steady state concentrations of metabolite 4 while keeping that of metabolite 3 above a prescribed minimum value). It is also worth mentioning, that one can replicate the above analysis to design a controller for branched metabolic networks with arbitrarily long main pathway and branch.



**Fig. 7.7** Global stability. **a** Concentration of metabolite 3 versus time. (**a Inset**) After a rapid transient, the initial metabolite concentrations become irrelevant; the metabolites quickly reach their quasi-steady state that depends exclusively on the value of the enzyme concentrations. **b** Concentration of the branch enzyme versus time. Four trajectories with different initial conditions are plotted alongside the equilibrium (*black, dashed*). All trajectories converge to the equilibrium. If the initial enzyme concentration is higher than its equilibrium value (as it is the case for the *dark blue* and *green* trajectories), the branch drains resources away from the native pathway depleting the concentration of metabolite 2 and, as a consequence, that of metabolite 3 too. The drop in concentration of 3 is detected by the genetic controller and the expression of the branch enzyme is repressed. This causes the enzyme concentration to return back to its equilibrium level, and that of metabolites 2 and 3 to return back to theirs

## 7.4 Discussion

In this chapter, we exploited the discrepancy in the speeds at which metabolic reactions and gene expression occur to justify the reduction of genetic-metabolic networks via time scale separation. If applicable, time scale separation reduces a model with  $n$  ‘fast’ variables and  $m$  ‘slow’ variables to one with just the  $m$  ‘slow’ variables. Such a model reduction can have strong benefits with regards to obtaining analytical results on the model (e.g., see [2, 16]).

The framework we use to describe genetic-metabolic systems is flexible. The assumptions made on the enzyme kinetics are minimal and are satisfied by a wide collection of kinetics models employed in the literature [3]. Furthermore, we make few assumptions regarding the ODEs describing the enzyme dynamics. Thus, we allow for a wide range of models for enzyme expression, with the notable exception of switch like models occasionally employed, e.g., [16]. However, our framework has some important drawbacks that can limit the applicability of Theorem 1.

First, we deal only with enzymatic reactions, i.e., reactions catalysed by an enzyme. Although many reactions involved in cellular metabolism are enzymatic reactions [3], there are also some that are not. This is not too hard to overcome; if non-enzymatic reactions are included in the network, then, following an approach nearly identical to that discussed in this chapter, one can obtain similar results regarding the validity of time scale separation.

Implicit in our framework is the assumption that each reaction has a single reactant. One could potentially include reactions with multiple reactants by following the

example set by Jackson, Horn and Feinberg and in their work on chemical reaction network theory (CNRT) [5]. They introduce the idea of chemical complexes, separate from chemical species (what we refer to as ‘metabolites’). For example, if one has the reaction  $A + 2B \rightarrow C$ ,  $A$ ,  $B$  and  $C$  are the chemical species involved in the reaction and  $A + 2B$  and  $C$  are the chemical complexes.

Another subtle but important issue is that the enzyme kinetics our framework is aimed for (e.g., Michaelis Menten, Hill type functions, etc.) are, themselves, the outcome of a previous reduction involving a quasi-steady state approximation. Key to these reductions is the assumption that the enzyme concentrations are constant. Although this is not the case in the type of models we are examining, where the enzyme concentrations are modelled as dynamic variables, there has recently been some progress in showing that these reductions are also valid if the enzyme concentrations vary, see [10].

The applicability of our results to the class of genetic-metabolic systems we consider has two main limitations. The first is that to carry out the reduction, one must show that the premise of Theorem 1 is satisfied. The second is that our results are only applicable to acyclic networks, i.e., networks that satisfy Condition 1. The former is not as much of a hindrance as one expects it to be; the enzyme dynamics, often, are such that the premise of Theorem 1 is not hard to satisfy. The latter is more serious, in particular because it rules out networks with reversible reactions. However, one can build on our current result to construct a more general one for the case of certain non-acyclic networks, e.g., ones that include reversible reactions.

To apply our result, one must first be able to find a compact subset of the set of all enzyme concentrations such that Condition 3 is satisfied, that is forward invariant with respect to the reduced model (7.14). Often, in models for enzyme dynamics, the differential equation describing the evolution of an individual enzyme is coupled to the metabolites and other enzymes via saturable functions [2, 14, 15]. Hence, one can often extract certain differential inequalities regarding the time evolution of individual enzymes that are decoupled from the other metabolites and enzymes. These can then be used to find the desired forward invariant regions. Indeed, this is exactly what we did in our example network, see Box 7.6.

The requirement that the network must be acyclic, i.e., that it satisfies Condition (1), is a limitation. This is especially true because it rules out networks with reversible reactions. However, if one is willing to impose some more conservative inequalities than those in Condition 3, it is straightforward to extend the result to a significantly more general class of networks.

Our proof for the acyclic case consists of showing that the metabolite dynamics,  $\dot{s} = f(s, e)$ , are such that the premise of Tikhonov’s Theorem (Theorem 2) is satisfied. In particular, we show that for any fixed enzyme vector  $e \in \mathbb{R}_{>0}^m$  the system  $\dot{s} = f(s, e)$  has globally asymptotically stable equilibrium. To do this, we use the fact that the network is acyclic to decompose the system  $\dot{s} = f(s, e)$  into a series of interconnected 1 dimensional subsystems, or blocks, such that the input the  $i$ th subsystem comes only from the previous  $i - 1$  systems. We then prove certain properties about these subsystems (essentially that they are *converging input converging state* (CICS)) and use these to establish properties about the complete

system required to satisfy the theorem's premise. However, there is no reason why to only use 1-dimensional subsystems other than that it is easier to show that they are CICS. If one can show that larger blocks, e.g., a 2-dimensional block representing a reversible reaction, are also CICS then the result would be almost immediate for 'block-acyclic' networks containing a mixture of 1 dimensional irreversible reaction blocks and larger blocks. Indeed, by imposing stronger inequalities than those in Condition 3, it is straightforward to show that much more general blocks are CICS, e.g., chains of reversible reactions and loops of irreversible reactions. However, to simplify this exposition we limit ourselves to the acyclic case. Strictly speaking, to satisfy the premise of Tikhonov's Theorem, one must also show that the eigenvalues of the Jacobian of  $\dot{s} = f(s, e)$  all have negative real parts. This can be done easily because the network being acyclic implies that the Jacobian is triangular. If one considers a block acyclic network then the Jacobian will be block triangular. All that one needs to show in this case is that the eigenvalues of the Jacobian of each of the blocks have negative real parts.

An interesting alternative would be to attempt to use the existing CNRT machinery, specifically the Deficiency Zero Theorem [5], to re-derive and potentially extend our results, at least to networks with mass-action kinetics.

**Acknowledgments** We thank Aivar Sootla for very useful discussions about various topics described in this chapter and Alexandros Houssein and Keshava Murthy for their valuable advice regarding how to improve this script.

## Appendix

In the appendices we assume that the reader has some familiarity with non-linear systems theory. Specifically, we assume that the reader is comfortable with the various notions of stability of equilibria, Lyapunov functions and the existence and uniqueness results. If not, we refer the reader to the excellent text [8].

We begin by presenting Tikhonov's Theorem over finite time intervals and some related results. Next, we discuss the notion of converging input converging state systems. Lastly, we employ the previous two to prove Lemmas 1 and 2 and Theorem 1.

Throughout the appendices we use  $\|\cdot\|$  to denote any vector norm.

### *A: Tikhonov's Theorem*

As discussed in the main text, a method for dimensionality reduction of non-linear systems is time scale separation. This is applicable to systems whose state variables exhibit large differences in the 'speed' of their time responses. Core to time scale separation is the following result first proved by Tikhonov 60 years ago [21, 22]. The version of it presented here is not the original version by Tikhonov, but instead the version published by Vasil'eva in 1963, which we find easier to work with.

**Theorem 2** [9, 23] *Let  $f: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $g: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$  both be smooth functions. Consider the system*

$$\varepsilon \dot{z}(t) = f(x(t), z(t)), \quad z(0) = z_0, \quad z \in \mathbb{R}^m, \quad (7.19a)$$

$$\dot{x}(t) = g(x(t), z(t)), \quad x(0) = x_0, \quad x \in \mathbb{R}^n, \quad (7.19b)$$

where  $\varepsilon > 0$ . Assume for all  $t \in [0, T]$  where  $T \in \mathbb{R}_{\geq 0}$  that (7.19) has the unique solutions  $x(t), z(t)$ . Consider the following conditions:

1. *There exists a unique function  $\phi(\cdot)$  such that  $g(\bar{x}(t), \phi(\bar{x}(t))) = 0$  for all  $t \in [0, T]$  where  $\bar{x}(t)$  denotes the unique solution over  $[0, T]$  of the reduced system  $\dot{\bar{x}} = g(\bar{x}, \phi(\bar{x}))$ ,  $x(0) = x_0$ .*
2. *Consider the ‘boundary layer’ system*

$$\frac{d\hat{z}}{d\tau}(\tau) = f(x_0, \hat{z}(\tau) + \phi(x_0)). \quad (7.20)$$

*Assume that the equilibrium  $\hat{z} = 0$  of (7.20) is globally asymptotically stable, uniformly in  $x_0$ .*

3. *The eigenvalues of  $\left[ \frac{\partial f}{\partial z}(\cdot) \right]$  evaluated along  $\bar{x}(t), \bar{z}(t)$ , have real parts smaller than a fixed negative number, i.e.,*

$$\operatorname{Re} \left( \lambda_i \left( \left[ \frac{\partial f}{\partial z} \right] (\bar{x}(t), \bar{z}(t)) \right) \right) \leq -c, \quad c \in \mathbb{R}_{>0}, \quad \forall i, \quad \forall t \geq 0.$$

*where  $\operatorname{Re}(a)$  denotes the real part of  $a \in \mathbb{C}$  and  $\lambda_i(A)$  denotes the  $i$ th eigenvalue of  $A \in \mathbb{R}^{n \times n}$ .*

*If the three conditions above are satisfied, then relations (7.21) and (7.22) hold for all  $t \in [0, T]$  and there exists a time  $t_1 \geq 0$ ,  $O(\varepsilon \ln(1/\varepsilon))$ , such that (7.23) holds for all  $t \in [t_1, T]$ .*

$$x(t) = \bar{x}(t) + O(\varepsilon). \quad (7.21)$$

$$z(t) = \bar{z}(t) + \hat{z}(t) + O(\varepsilon). \quad (7.22)$$

$$z(t) = \bar{z}(t) + O(\varepsilon). \quad (7.23)$$

The Theorem’s first condition ensures that there exists a well defined reduced model. The second condition verifies that, initially, the trajectory of the complete system rapidly converges to the one of the reduced system. The third condition guarantees that after the initial transient dies out the trajectory of the complete system remains close to the that of the reduced system. It is worth mentioning, that the above is Tikhonov’s theorem restricted to the special case when the systems are time invariant and (7.19a) has a unique root. For an excellent treatment of Tikhonov’s



theorem (including its most general form) and its applications in control theory see [9].

In verifying the theorem's last two conditions the following two lemmas will be useful.

**Lemma 3** [22] *Consider the boundary system (7.20). Assume that  $f$  and the root  $\phi$  are continuous functions and that  $x_0 \in \mathcal{P}$  where  $\mathcal{P}$  is a compact subset of  $\mathbb{R}^m$ . Suppose that for all  $x_0 \in \mathcal{P}$ , the origin of (7.20) is globally asymptotically stable. Then the origin of (7.20) is globally asymptotically stable, uniformly in  $x_0$ .*

**Lemma 4** *Consider  $f(\cdot)$  in (7.19). Let  $A$  be a compact subset of  $\mathbb{R}^{n+m}$  and suppose that*

$$\operatorname{Re} \left( \lambda_i \left( \left[ \frac{\partial f}{\partial z} \right] (x, z) \right) \right) < 0, \quad \forall i, \quad \forall [x, z]^T \in A.$$

*Then, there exists a  $c \in \mathbb{R}_{>0}$  such that*

$$\operatorname{Re} \left( \lambda_i \left( \left[ \frac{\partial f}{\partial z} \right] (x, z) \right) \right) \leq -c, \quad \forall \begin{bmatrix} x \\ z \end{bmatrix}^T \in A.$$

*Proof* First, we show that

$$\lambda^*(x, z) := \max_i \left( \lambda_i \left( \left[ \frac{\partial f}{\partial z} \right] (x, z) \right) \right), \quad (7.24)$$

that is, the maximum real part of the eigenvalues of the Jacobian, is a continuous function of  $x$  and  $z$ .

The eigenvalues are the roots of the characteristic polynomial of the Jacobian (i.e., the solutions to  $\det \left( \lambda I - \left[ \frac{\partial f}{\partial z} \right] (x, z) \right) = 0$  where  $\lambda \in \mathbb{C}$ ). The roots of a polynomial depend continuously on the coefficients of a polynomial. The coefficients of the characteristic polynomial of the Jacobian above depend continuously of the entries of the Jacobian. The entries of the Jacobian are continuous functions of  $x$  and  $z$ . The composition of two continuous functions is also a continuous function. Thus, the eigenvalues of the Jacobian are continuous functions of  $x$  and  $z$ . Thus, (7.24) is a continuous function of  $x$  and  $z$ .

The supremum of a continuous function over a compact set is achieved by an element in the set. This fact and the lemma's premise imply that  $\sup_{[x,z]^T \in A} \lambda^*(x, z) < 0$  which completes the proof.  $\square$

## ***B: Converging Input Converging State Systems***

In Appendix C, we need to prove that the unique equilibrium of the network with the enzyme concentrations fixed in time (system (7.8)) is globally asymptotically

stable (GAS). To accomplish this we exploit the acyclicity of the network to break system (7.8) down into  $n$  one dimensional subsystems and study how they interact. To this end, we introduce the notions of *converging input bounded state* (CIBS) and *converging input converging state* (CICS) systems. These were originally presented in [20] and relate to other more well known concepts such as *input to state stable* (ISS) systems.

**Definition 1** We say that  $u(\cdot)$  is an *input* if it is a continuous function that maps from  $\mathbb{R}_{\geq 0}$  to  $\mathbb{R}^m$ .

Now, consider the non-autonomous system

$$\dot{x}(t) = f(x(t), u(t)), \quad (7.25)$$

where  $f(\cdot)$  is continuous,  $x \in \mathbb{R}^n$  and  $u(\cdot)$  is an input. In addition, consider the same system with ‘zero input’

$$\dot{x}(t) = f(x(t), 0). \quad (7.26)$$

**Definition 2** System (7.25) is said to be *converging input bounded state* (CIBS) if for any input  $u(\cdot)$  such that  $u(t) \rightarrow 0$  as  $t \rightarrow +\infty$  and for any initial conditions  $x_0 \in \mathbb{R}^n$ , the solution exists for all  $t \geq 0$  and is bounded.

**Definition 3** System (7.25) is said to be *converging input converging state* (CICS) if for any input  $u(\cdot)$  such that  $u(t) \rightarrow 0$  as  $t \rightarrow +\infty$  and for any initial conditions  $x_0 \in \mathbb{R}^n$ , the solution exists for all  $t \geq 0$  and converges to 0 as time tends to infinity.

**Lemma 5** Assume that for any input,  $x(t)$  exists for all  $t \geq 0$ . Let  $V: \mathbb{R}^n \rightarrow \mathbb{R}$  be a continuously differentiable, bounded from below and radially unbounded (i.e.,  $\|x\| \rightarrow +\infty \Rightarrow V(x) \rightarrow +\infty$ ) function. If there exists constants  $\alpha > 0$  and  $\beta > 0$  such that

$$\dot{V}(x) = \frac{\partial V}{\partial x} f(x, u) \leq 0 \quad \forall (x, u) \in \mathbb{R}^{n+m}: \|x\| \geq \beta, \|u\| \leq \alpha,$$

then system (7.25) is CIBS.

*Proof* We prove by contradiction. Assume that the premise of the lemma is satisfied and that there exists a  $u(t)$  such that  $\|u(t)\| \rightarrow 0$  as  $t \rightarrow +\infty$  but  $x(t)$  is unbounded. By our premise,  $x(t)$  is defined for all  $t \geq 0$ . Thus, there does not exist a finite escape time, i.e., there does not exist a time  $T \geq 0$  such that  $\|x(t)\| \rightarrow +\infty$  as  $t \rightarrow T$ . Thus, the fact that  $x(t)$  is unbounded implies that  $\|x(t)\| \rightarrow +\infty$  as  $t \rightarrow +\infty$ .

Now,  $\|u(t)\| \rightarrow 0$  as  $t \rightarrow +\infty$  implies that there exists a  $t_1 \geq 0$  such that  $\forall t \geq t_1, \|u(t)\| \leq \alpha$ . In addition,  $\|x(t)\| \rightarrow +\infty$  as  $t \rightarrow +\infty$  implies that there exists a  $t_2 \geq 0$  such that  $\forall t \geq t_2, \|x(t)\| \geq \beta$ . Let  $t_3 := \max\{t_1, t_2\}$ . Thus,  $\forall t \geq t_3, \dot{V}(x(t)) \leq 0$  which implies that  $\forall t \geq t_3, V(x(t)) \leq V(x(t_3))$ . This implies that  $x(t)$  does not tend to  $+\infty$  as  $t$  tends to  $+\infty$ . We have reached a contradiction.  $\square$

**Theorem 3** [20] *If 0 is a GAS equilibrium of (7.26) then CIBS and CICS are equivalent for (7.25).*

**Theorem 4** [20] *Consider the cascade formed by system (7.25) and the autonomous system  $\dot{y} = g(y)$ ,*

$$\dot{x} = f(x, y), \quad (7.27a)$$

$$\dot{y} = g(y), \quad (7.27b)$$

where  $g$  is continuous,  $y \in \mathbb{R}^m$ . Assume the origin of (7.27b) is GAS and that (7.25) is CICS. Then the origin of (7.27) is GAS.

### C: Proof of the Main Results

We begin by demonstrating a series of results regarding the metabolic model when enzymes are kept at a fixed value. In other words, up to and including the proof of Lemma 1 we neglect the enzyme dynamics (7.1b) and assume  $e(t) \equiv e$ , where  $e \in \mathbb{R}_{>0}^m$  is a constant such that Conditions 1–3 hold. In Sect. 7.3.1, we argued that if Conditions 1–3 are satisfied, the metabolic network (7.8) has a unique equilibrium  $\bar{s}$ .

We now establish global asymptotic stability of the equilibrium. To do this, instead of studying the behaviour of the whole network in one go, we examine the behaviour of individual metabolites, or individual *subsystems* first, and then using these we establish the stability property for the whole network. We call

$$\dot{x}(t) = f_1(x(t), e), \quad x(0) = x_0 \in \mathbb{R}_{\geq 0}$$

the 1st *subsystem* where  $f_1$  is defined as in (7.8). Similarly, we call

$$\dot{x}(t) = f_i(w(t), x(t), e), \quad x(0) = x_0 \in \mathbb{R}_{\geq 0}$$

the  $i$ th *subsystem*<sup>1</sup> where  $w: \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}^{i-1}$  plays the role of an input and  $f_i$  is defined as in (7.8) for  $i = 2, \dots, n$ . Note that, given that the domain of  $f_i$ , with  $i = 2, \dots, n$ , is  $\mathbb{R}_{\geq 0}^{i-1} \times \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$  (the reaction rates,  $v_{j \rightarrow i}$  are only defined for non-negative arguments), it is important that the range of  $w$  is  $\mathbb{R}_{\geq 0}^{i-1}$  instead of  $\mathbb{R}^{i-1}$ . For this reason, if we want to employ the CICS machinery introduced in Appendix 2, we first must alter slightly our definition of an input  $u(\cdot)$  (Definition 1, Appendix B).

**Definition 4** We say that  $u(\cdot)$  is an input to the system  $\dot{x} = f(x, u)$ ,  $f: A \times B \rightarrow \mathbb{R}^n$  where  $A \times B \subset \mathbb{R}^n \times \mathbb{R}^m$ , if it is a continuous function that maps from  $\mathbb{R}_{\geq 0}$  to  $B$ .

<sup>1</sup> Here we are abusing slightly our notation by writing the first  $i - 1$  scalar arguments of  $f_i$  as a single  $i - 1$  dimensional vector argument.

It can be shown, in a similar manner as in Appendix B and [20], that Lemma 5 and Theorems 3 and 4 hold if one replaces the original definition of an input (Definition 1) with the one above (Definition 4) and  $x_0 \in \mathbb{R}^n$  with  $x_0 \in A$ .

Returning to our original problem, it is convenient to introduce the change of coordinates  $z := x - \bar{s}$  and  $u(\cdot) := w(\cdot) - \bar{s}^i$  where  $\bar{s}^i := [\bar{s}_1 \cdots \bar{s}_{i-1}]^T$  for  $i = 2, \dots, n$ . Then, we can re-write the 1st subsystem as

$$\dot{z}(t) = f_1(z(t) + \bar{s}_1, e), \quad z(0) = z_0 \in [-\bar{s}_1, +\infty).$$

and the  $i$ th subsystem

$$\dot{z}(t) = f_i(u(t) + \bar{s}^i, z(t) + \bar{s}_i, e), \quad z(0) = z_0 \in [-\bar{s}_i, +\infty). \quad (7.28)$$

for  $i = 2, \dots, n$ . In addition, from now onwards we will say an input  $u(\cdot)$  meaning an input to the  $i$ th subsystem (7.28) in the sense of Definition 4.

**Proposition 1** *For any input given  $u(\cdot)$ , then the  $i$ th subsystem has a unique, continuous solution  $z(t) \in [-s_i, +\infty)$  for all  $t \geq 0$ .*

*Proof* Each component of  $f(\cdot)$  is a linear combination of globally Lipschitz continuous functions (Assumptions 1 and 2), hence  $f(\cdot)$  is globally Lipschitz continuous as well. This and the definition of  $u(\cdot)$  (which implies that it is a continuous function of  $t$ ), ensure that the  $i$ th subsystem,  $\dot{z} = f_i(u(t) + \bar{s}^i, z(t) + \bar{s}_i, e)$ , satisfies the usual conditions for global existence of solutions of time varying systems. Hence the  $i$ th subsystem has a unique, continuous solution  $z(t)$  that exists for all  $t \geq 0$ . Then, due to the positive definiteness of the  $g_i$ s and  $h_i$ s

$$z = -\bar{s}_1 \Rightarrow \dot{z} = f_1(0, e) = I_1 - h_1(0, e) = I_1 \geq 0$$

which proves that  $z(t) \in [-\bar{s}_1, +\infty)$  for all  $t \geq 0$  were  $z(t)$  is the solution of the 1<sup>st</sup> subsystem, and

$$\begin{aligned} z = -\bar{s}_i \Rightarrow \dot{z} &= f_i(u(t) + \bar{s}^i, 0, e) \\ &= g_i(u(t) + \bar{s}^i, e) - h_i(0, e) = g_i(u(t) + \bar{s}^i, e) \geq 0 \end{aligned}$$

which proves that  $z(t) \in [-\bar{s}_i, +\infty)$  for all  $t \geq 0$  were  $z(t)$  is the solution of the  $i$ th subsystem,  $i = 2, \dots, n$ .  $\square$

Proposition 1 is important for two reasons. First, it allows us to regard the state space of  $i$ th subsystem, (7.28), to be  $[-\bar{s}_i, +\infty)$  instead of  $\mathbb{R}$ . This makes sense, we are only interested in non-negative concentrations of the metabolites. Second, it shows that the vector containing the state of the first  $i - 1$  subsystems is input to the  $i$ th subsystem, in the sense of Definition 4.

**Proposition 2** *The  $i$ th subsystem is CIBS, for any  $i = 2, \dots, n$ .*

*Proof* Let  $V : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$  be defined as

$$V(z) := \frac{1}{2}z^2 \Rightarrow \dot{V}(z) = \frac{\partial V}{\partial z} \dot{z} = z\dot{z} = z \left( g_i(u + \bar{s}^i, e) - h_i(z + \bar{s}_i, e) \right).$$

By Condition 3,  $\hat{h}_i(e) > g_i(\bar{s}^i, e)$  thus  $\hat{h}_i(e) \geq g_i(\bar{s}^i, e) + \delta_1$ , for some  $\delta_1 > 0$ . In addition, by continuity and monotonicity of  $g_i$  (monotonicity in each of its arguments), there exists a sufficiently small  $\alpha > 0$  such that

$$g_i(\alpha \mathbb{1} + \bar{s}^i, e) - g_i(\bar{s}^i, e) \leq \frac{\delta_1}{2},$$

where  $\mathbb{1} := [1 \dots 1]^T$ . In addition,

$$\|u\|_\infty \leq \alpha \Rightarrow g_i(u + \bar{s}^i, e) \leq g_i(\alpha \mathbb{1} + \bar{s}^i, e) \leq g_i(\bar{s}^i, e) + \frac{\delta_1}{2} \leq \hat{h}_i(e) - \frac{\delta_1}{2}.$$

Hence, we have

$$\|u\|_\infty \leq \alpha \Rightarrow g_i(u + \bar{s}^i, e) - h_i(z + \bar{s}_i, e) \leq \hat{h}_i(e) - \frac{\delta_1}{2} - h_i(z + \bar{s}_i, e).$$

Because  $h_i(z + \bar{s}_i, e) \rightarrow \hat{h}_i(e)$  from below as  $z \rightarrow +\infty$  we can always find a  $\beta_1$  such that  $z \geq \beta_1 \Rightarrow \hat{h}_i(e) - h_i(z + \bar{s}_i, e) \leq \delta_2$  for any given  $\delta_2 > 0$ . In addition, because  $z \in [-\bar{s}_i, +\infty)$ ,  $\|z\| > \bar{s}_i$  implies  $\|z\| = z$ . Hence, choosing  $\delta_2 \leq \frac{\delta_1}{2}$  and defining  $\beta := \max\{\beta_1, \bar{s}_i + \varepsilon\}$ , where  $\varepsilon > 0$ , we have

$$u, z: \|u\|_\infty \leq \alpha, \|z\| \geq \beta \Rightarrow \dot{V}(z) \leq z(\hat{h}_i(e) - h(\hat{s}_i, e) - \frac{\delta_1}{2}) \leq z(\delta_2 - \frac{\delta_1}{2}) \leq 0$$

Then, applying Lemma 5 completes the proof.  $\square$

**Proposition 3** *The origin of  $i$ th subsystem with zero input (i.e.,  $u(t) \equiv 0$ ) is a globally asymptotically stable equilibrium, for any  $i = 1, \dots, n$ .*

*Proof* We use the Lyapunov function

$$\begin{aligned} V(z) &:= \frac{1}{2}z^2 \Rightarrow \dot{V}(z) = \frac{\partial V}{\partial z} \dot{z} = z f_i(\bar{s}^i, z(t) + \bar{s}_i, e) \\ &= z(g_i(\bar{s}_1, \dots, \bar{s}_{i-1}, e) - h_i(z + \bar{s}_i, e)). \end{aligned}$$

By the definition of  $\bar{s}$ , we have that  $g_i(\bar{s}_1, \dots, \bar{s}_{i-1}, e) = h_i(\bar{s}_i, e)$ . So

$$\dot{V}(z) = z(h_i(\bar{s}_i, e) - h_i(z + \bar{s}_i, e)).$$

Due to the strict monotonicity of  $h_i, z$  and  $(h_i(\bar{s}_i, e) - h_i(z + \bar{s}_i, e))$  have opposite signs and are both equal to zero if and only if  $z = 0$ . Hence, applying Lyapunov's Direct Method completes the proof.  $\square$

**Proposition 4** *The  $i$ th subsystem is CICS, for any  $i = 1, \dots, n$ .*

*Proof* This follows directly from Propositions 2 and 3 and Theorem 3.  $\square$

With these preliminary results in mind, we are now ready to prove Lemma 1.

*Proof* (Lemma 1) As previously pointed out, the solution to the first subsystem is an input to the second subsystem, in the sense of Definition 4. Consider the cascade obtained by setting the input of the 2nd subsystem to the state of the 1st subsystem,

$$\begin{aligned}\dot{z}_1(t) &= f_1(z_1(t) + \bar{s}_1, e), \\ \dot{z}_2(t) &= f_2(z_1(t) + \bar{s}_1, z_2(t) + \bar{s}_2, e).\end{aligned}$$

Propositions 3 (i.e., the origin of the 1st subsystem is a GAS equilibrium) and 4 (i.e., the 2nd subsystem is CICS) and Theorem 4 (i.e., that the origin of the interconnection of an autonomous system which has a GAS equilibrium at the origin and a CICS system has a GAS equilibrium) imply that the origin of the above cascade is GAS. Then, by induction, we see that the origin of the system obtained by iteratively cascading the  $i$ th subsystem with the cascade formed by the previous  $i - 1$  subsystems is a GAS equilibrium. In other words, the origin of

$$\dot{z} = f(z + \bar{s}, e)$$

is a GAS equilibrium, which completes the proof.  $\square$

**Proposition 5** *Let  $A$  denote the subset of  $\mathbb{R}_{>0}^m$  whose elements are such that Condition 3 holds. There exists a unique function  $\phi: A \rightarrow \mathbb{R}_{>0}^n$  such that  $f(\phi(e), e) = 0$  for all  $e \in A$ . Furthermore, this function is continuously differentiable and globally Lipschitz continuous.*

*Proof* Existence and uniqueness of  $\phi$  follows from our discussion in Sect. 7.3.1 of the main text regarding the existence and uniqueness of an equilibrium if the enzymes are constant. Each component of  $f(\cdot)$  is a linear combination of continuously differentiable and globally Lipschitz continuous functions (Assumptions 1 and 2). Thus,  $f(\cdot)$  is continuously differentiable and globally Lipschitz continuous or, equivalently its partial derivatives exists everywhere, are continuous and bounded. The fact that  $f(\phi(e), e) = 0$  for all  $e \in A$  implies that the total derivative of  $f(\cdot)$  along  $[\phi(e) \ e]^T$  is also equal to zero, i.e.,  $f'(\phi(e), e) = 0$  for all  $e \in A$ . The total derivative of a function exists and is continuous if and only if the partial derivatives of the function exist and are continuous. Hence,

$$\frac{\partial f}{\partial \phi}(\phi(e), e) \frac{\partial \phi}{\partial e}(e) + \frac{\partial f}{\partial e}(\phi(e), e) = 0$$

which implies that

$$\frac{\partial \phi}{\partial e}(e) = - \left( \frac{\partial f}{\partial \phi}(\phi(e), e) \right)^{-1} \frac{\partial f}{\partial e}(\phi(e), e).$$

By Condition 1,  $v_{j \rightarrow i} \equiv 0$  if  $i < j$ . Hence,  $i < j \Rightarrow \frac{\partial f_i}{\partial \phi_j}(\phi(e), e) = \frac{\partial v_{j \rightarrow i}}{\partial \phi_j}(\phi_j(e), e) = 0$ . Thus,  $\frac{\partial f}{\partial \phi}(\phi(e), e)$  is lower triangular. Furthermore, by Condition 2,  $h_i$  is strictly increasing, hence

$$\frac{\partial f_i}{\partial \phi_i}(\phi(e), e) = - \frac{\partial h_i}{\partial \phi_i}(\phi_i(e), e) < 0.$$

Thus,  $\left( \frac{\partial f}{\partial \phi}(\phi(e), e) \right)^{-1}$  exists for all  $e \in A$ . Hence,  $\frac{\partial \phi}{\partial e}(e)$  exists for all  $e \in A$ . Furthermore,  $\frac{\partial \phi}{\partial e}(e)$  is continuous and bounded which shows that  $\phi$  is continuously differentiable and globally Lipschitz continuous.  $\square$

We are now in a position to prove Lemma 2 and Theorem 1.

*Proof* (Lemma 2) The existence and uniqueness of  $s(t)$  and  $e(t)$  follow from our assumption that  $f(\cdot)$  and  $g(\cdot)$  are smooth and globally Lipschitz continuous (Assumptions 1–3). The existence and uniqueness of  $\phi(\cdot)$  is proven in Proposition 5. The domain of  $\phi(\cdot)$  is  $A$ . Thus, (7.14) is well-defined if and only if  $\bar{e}(t)$  remains in  $A$ . This is ensured by the premise,  $B \subseteq A$  is forward invariant with respect to (7.14) and  $e_0 \in B$ . In addition,  $g(\cdot)$  and  $\phi(\cdot)$  are globally Lipschitz continuous (Assumption 3, Proposition 5, respectively), which implies that (7.14) satisfies the usual conditions for global existence and uniqueness solutions.  $\square$

*Proof* (Theorem 1) The proof is an application of Tikhonov’s Theorem on finite time intervals (Theorem 2). The existence and uniqueness of  $\phi(\cdot)$  satisfies the first condition in the premise of Theorem 2 which requires that the metabolite dynamics,  $f(s, e)$ , has a unique root.

The second condition of Tikhonov’s Theorem is that  $z = \phi(e_0)$  is a globally asymptotically stable equilibrium, uniformly in  $e_0$ , of the boundary layer system  $\dot{z} = f(z, e_0)$ . Lemma 1 shows that for any given  $e_0 \in B \subseteq A$ ,  $\phi(e_0)$  is a globally asymptotically stable equilibrium of  $\dot{z} = f(z, e_0)$ . The fact that  $B$  is compact combined with the previous statement form the premise of Lemma 3. Then, Lemma 3 establishes the desired result, i.e., that the equilibrium  $z = \phi(e_0)$  is a globally asymptotically stable, uniformly in  $e_0$ .

Proposition 5 shows that  $\phi(\cdot)$  is continuous. Because  $\bar{e}(t) \in B$  for all time, and  $B$  is a compact set,  $\bar{s}(t) = \phi(\bar{e}(t))$  must also be confined to some compact set. In the proof of Proposition 5 we established that for any given  $e \in B \subseteq A$ , the eigenvalues of the Jacobian of the boundary layer system evaluated at  $[\phi(e), e]^T$ ,  $\frac{\partial f}{\partial \phi}(\phi(e), e)$ , have negative real parts. The previous two statements form the premise of Lemma 4 which shows that the eigenvalues of the Jacobian of the boundary layer system,

evaluated along  $[\phi(\bar{e}(t)) \bar{e}(t)]^T$  have real parts smaller than a certain negative real number, i.e., that the third condition of Tikhonov's theorem is satisfied.  $\square$

## References

1. Alon U (2006) An introduction to systems biology: design principles of biological circuits. Chapman and Hall/CRC, London
2. Baldazzi V, Ropers D, Geiselmann J, Kahn D, de Jong H (2012) Importance of metabolic coupling for the dynamics of gene expression following a diauxic shift in *Escherichia coli*. *J Theor Biol* 295:100–115
3. Cornish-Bowden A (2004) Fundamentals of enzyme kinetics, 3rd edn. Portland Press, London
4. Craciun G, Pantea C, Sontag ED (2011) Graph theoretical analysis of multistability and monotonicity for biochemical reaction networks, vol 4. Springer, Berlin, pp 63–72
5. Feinberg M (1987) Chemical reaction network structure and the stability of complex isothermal reactors I. The deficiency zero and deficiency one theorems. *Chem Eng Sci* 42(10):2229–2268
6. Flach EH, Schnell S (2006) Use and abuse of the quasi-steady-state approximation. *Syst Biol* 153(4):187–191
7. Heinrich R, Schuster S (1996) The regulation of cellular systems. Springer, Berlin
8. Khalil HK (2002) Nonlinear systems, 2nd edn. Prentice Hall, Englewood Cliffs
9. Kokotovic P, Khalil HK, O'Reilly J (1986) Singular perturbation methods in control: analysis and design. Academic Press, New York
10. Kumar A, Josić K (2011) Reduced models of networks of coupled enzymatic reactions. *J Theor Biol* 278(1):87–106, 1101.1104
11. Lin CC, Segel LA (1988) Mathematics applied to deterministic problems in the natural sciences. SIAM, Philadelphia
12. Madigan MT, Martinko JM, Stahl DA, Clark DP (2011) Brock biology of microorganisms, 13th edn. Pearson Education, New Jersey
13. Nielsen J, Keasling JD (2011) Synergies between synthetic biology and metabolic engineering. *Nat Biotechnol* 29(8):693–695
14. Oyarzún DA, Stan GB (2012) Design tradeoffs in a synthetic gene control circuit for metabolic networks. In: Proceedings of the 31st American control conference, Montreal
15. Oyarzún DA, Stan GB (2013) Synthetic gene circuits for metabolic control: design tradeoffs and constraints. *J Royal Soc Interface* 10(78):20120671
16. Oyarzún DA, Chaves M, Hoff-Hoffmeyer-Zlotnik M (2012) Multistability and oscillations in genetic control of metabolism. *J Theor Biol* 295:139–153
17. Radde N, Bar NS, Banaji M (2010) Graphical methods for analysing feedback in biological networks. A survey. *Int J Syst Sci* 41(1):35–46
18. Scheer M, Grote A, Chang A, Schomburg I, Munaretto C, Rother M, Sohngen C, Stelzer M, Thiele J, Schomburg D (2011) Brenda, the enzyme information system in 2011. *Nucleic Acids Res* 39:D670–D676
19. Segel LA, Slemrod M (1989) The quasi-steady-state assumption: a case study in perturbation. *SIAM Rev* 31:446–477
20. Sontag ED (1989) Remarks on stabilization and input-to-state stability. In: Proceedings of the 28th IEEE conference on decision and control. IEEE, pp 1376–1378
21. Tikhonov AN (1948) On dependence of the solutions of differential equations on a small parameter. *Mat Sb* 22:193–204 (in Russian)
22. Tikhonov AN (1952) Systems of differential equations containing a small parameter multiplying the derivative. *Mat Sb* 31:575–586 (in Russian)
23. Vasil'eva AB (1963) Asymptotic behaviour of solutions to certain problems involving non-linear differential equations containing a small parameter multiplying the highest derivatives. *Uspekhi Mat Nauk* 18:15–86



24. Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U (2004) Just-in-time transcription program in metabolic pathways. *Nat Genet* 36(5):486–491
25. Zhang F, Keasling J (2011) Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol* 19(7):323–329

# Chapter 8

## Networks, Metrics, and Systems Biology

Soumen Roy

**Abstract** The theory of complex networks plays an important role in Systems Biology. There are extensive discussions in literature about biological networks bearing the knowledge of function and possessing the key to “emergent properties” of the system. One would naturally assume that many network metrics need to be thoroughly studied to extract maximum information about the system. Interestingly however, most network papers discuss at most two three metrics at a time. What justifies the choice of a few metrics, in place of a comprehensive suite of network metrics? Is there any scientific basis of the choice of metrics or are they invariably handpicked? More importantly, do these few handpicked metrics carry the maximum information extractable about the biological system? This chapter discusses how any why the study of multiple metrics is necessary in biological networks and systems biology.

**Keywords** Complex networks · Steady state · Flux balance analysis (FBA) · Minimization of metabolic adjustment (MOMA) · Elementary mode analysis (EMA) · Topology · Topological analysis

### 8.1 Introduction

Modern high-throughput era has launched a flood of biological data. Apart from the obvious technical challenges of how to store and manage such copious amounts of data is, of course, the no less important challenge on how to interpret meaningful patterns in this flood of data. The manner in which biological interactions need to be mapped necessitates a separate language for its study. The theory of Complex Networks [1, 2], provides a good framework for scripting such interactions in modern biology. Without such a framework, we would not be able to ask questions about

---

S. Roy (✉)  
Bose Institute, 93/1 Acharya Prafulla Chandra Roy Road,  
Kolkata 700009, India  
e-mail: soumen@jcbosc.ac.in

the “emergent properties” and “systemic behavior” of complex biological systems under study.

It was not until the recent developments of complex networks that we had a mechanism for distinguishing or classifying different networks. Prior to this mathematicians, electrical engineers and computer scientists had made a very thorough study of random graphs. In particular, the models proposed by Erdos and Renyi had been the subject of extensive research. Modern developments in network theory [1, 2] showed that diverse networks drawn over all forms of life shared some common properties which could be quantified by the means of various network metrics. These are discussed in Sect. 8.2.

It was not only in the identification and classification of topological properties that these new findings were important. Much more useful insights were to follow. Among these, a very striking observation was that network topology possesses the potential of being a major determinant of biological function (or dysfunction). Relations between topological properties of network nodes (genes, proteins) and functional essentiality were uncovered in interaction networks [3, 4].

Long before the advent of the complex networks era, extensive modeling had been undertaken using steady-state flux balance approaches in metabolic networks [5] via methods like Flux Balance Analysis (FBA) [6] Minimization of Metabolic Adjustment (MOMA) [7] and Elementary Mode Analysis (EMA) [8].

However, even then, topological analysis has often yielded novel and valuable insight in metabolic networks. New parameters like synthetic accessibility have demonstrated sufficient promise in predicting the viability of knockout strains with accuracy comparable to approaches using biochemical parameters (like FBA etc.) on large, unbiased mutant data sets [9]. This is notable since determining synthetic accessibility does not require the knowledge of stoichiometry or maximal uptake rates for metabolic and transport reactions. On the other hand such knowledge is essential in FBA, MOMA and EMA. Interestingly, synthetic accessibility can be rapidly computed for a given network and has no adjustable parameters.

There are extensive discussions in literature about biological networks bearing the knowledge of biological function and possessing the key to “emergent properties” of the system. One would naturally assume thorough study of many network metrics would convey maximum information about the system. Interestingly however, most network papers discuss at most two three metrics at a time. What justifies the choice of a few metrics, in place of a comprehensive suite of network metrics? Is there any scientific basis of the choice of the metrics or are they invariably handpicked? More importantly, do these few handpicked metrics carry the maximum information extractable about the biological system? In the next few sections, we will attempt to answer these questions.

## 8.2 Network Metrics

In order to familiarize the readers who is uninitiated with the various network metrics, we provide at first very basic introduction in this section. In the next section, we then proceed to investigate how and why multiple network metrics can give us useful information.

The most common topological metric in networks is degree, which is henceforth denoted by  $k$ . It is the number of connections a node has to other nodes in the network and perhaps also with itself. The distribution of degree is perhaps the most well-studied item for almost all network systems. It is well-known that Erdos-Renyi Networks have a Poisson Degree Distribution. However, most real world networks, including biological networks have a heavy-tailed degree distributions. The most prominent feature of all heavy-tailed degree distributions is the presence of a few hubs or high degree nodes in a network with the simultaneous presence of leaves or low degree nodes.

Of special interest is a class of degree distributions which obey a power law

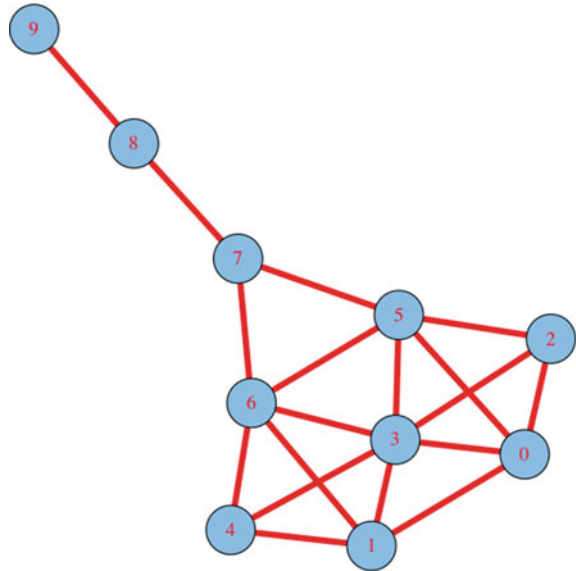
$$P(k) = k^{-\gamma} \quad (8.1)$$

It is obvious from the above equation that if  $k$  is replaced by  $ak$ , the form of the distribution remains invariant. Hence these distributions are also known as “scale-free” distributions. Power-laws have a special importance across the sciences, eg. in phase transitions, turbulence, Gutenberg-Richter law, Pareto Law, Zipf law etc. It is perhaps as a result of this strong presence of power laws across the sciences that many a times in recent literature one would almost invariably find papers classifying heavy tailed degree-distributions as “scale-free networks”. This is vexing since many degree distributions could be fit equally well or perhaps even better by other distributions. More so, because a very reliable statistical machinery for proper identification of scale-free networks has existed for quite some time [10].

The rather excessive mention of “hubs” in literature probably arises from the apparent conclusion that removal of such high-degree nodes could cause massive damage to the network. However, it has been clearly demonstrated by means of the “S-metric”, that even in networks with scale-free degree distributions, extremely important networks like the internet could well be structurally robust and functionally stable [11]. Therefore, plucking out the hubs from a network might not necessarily lead to a catastrophe as one might think at the first instance. On the contrary, the effect could be merely local.

That the hubs are not always the most important nodes in a network has been known for a long time. Social scientists have clearly demonstrated this via the analysis of graphs like the “Krackhardt kite graph” shown in Fig. 8.1 [12]. One of the most important properties of a network node which reflects this fact is known as “betweenness centrality”, [13]. It measures the fraction of all shortest paths which passes through a node. The role of betweenness is depicted in Fig. 8.1. At the first glance, it may appear that targeting a ‘hub’ (like 3) would cause intense damage to the network. However, the fact is that targeting a high-betweenness node like 7

**Fig. 8.1** The role of betweenness: it has been known for long that targeting high-betweenness node like 7 over hubs (like 3) could cause much more damage to this network



which obviously does not possess a relatively high degree would cause much more damage to this network.

In the world air transportation network, one would probably anticipate that most shortest flights between any two airports are likely to pass through cities like London and New York. However, actual analyses showed that 60 % of the 25 most connected airports of the world did not lie on many of these shortest paths. Instead airports like Anchorage and Port Moresby [14] lay on many of these shortest paths. The analysis on weighted networks might however change the results somewhat but the general importance of these results is not wasted. It is also known that for the US airline network, maximum damage would be done if the airports are targeted by betweenness rather than hubness [15]. Many papers using biological networks have found important results using betweenness [16–20]. Very recently the issue of controllability of complex biological networks has become very important [21]. The role of centrality metrics over degree is increasingly being discussed in this regard [22].

While any two nodes in a network might be connected by many paths, of special interest always, is the length of the shortest path between a pair of nodes. This is known as the *geodesic distance* between two nodes. There could of course more than one shortest path, each equal to the length in the network, connecting a pair of nodes. The *diameter* of the graph is another common metric which measures the longest such shortest path in the network.

Assortative mixing [23], quantifies the likeness of connections, i.e. whether high-degree nodes are *predominantly* connected to other high-degree nodes in a network or to low-degree nodes in a network. The degree assortativity, is defined as the Pearson correlation coefficient between the degrees of all pairs of connected vertices in the network. It should however be mentioned that being a correlation coefficient, assortativity is likely to have its limitations when “outliers” are present in the network. In these situations, Gini coefficient [24, 25]

$$G(k) = \frac{\sum_{i=1}^N \sum_{j=1}^N |k_i - k_j|}{N^2 \langle k \rangle}, \quad (8.2)$$

can be expected to capture the true picture, much better [15].

Assortativity has been studied in the context of various biological networks [26, 27] While it was earlier thought that all biological networks are disassortative, it has been subsequently found that protein contact networks could be assortative [26].

Another important metric in networks is clustering coefficient,

$$C = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}} \quad (8.3)$$

where a connected triple means a single vertex with edges running to an unordered pair of other vertices.

Another definition of the clustering coefficient, which has been given by Watts and Strogatz, who proposed a definition for the clustering coefficient of every node.

$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i} \quad (8.4)$$

For vertices with degree 0 or 1, the numerator and denominator are both zero, and we put  $C_i = 0$ . Then the clustering coefficient for the whole network is the average of the individual clustering coefficients of all nodes.

$$C = \frac{1}{n} \sum_i C_i \quad (8.5)$$

The Clustering coefficient of real world networks is almost invariably a few order of magnitudes higher than a random network formed of the same nodes and edges.

Rich club coefficient [28, 29],  $\phi(k) = 2E_k/N_k(N_k - 1)$  is the ratio, for every degree  $k$ , of the number of actual to the number of potential edges for nodes with degree greater than  $k$ ; where  $N_k$  is the number of nodes with degree larger than  $k$ , and  $E_k$  is the number of edges among those nodes. The human brain which is a complex network of interlinked regions displays a rich-club organization [30].

Also important network formulations like spectral graph theory are also known to shed valuable insights in graphs and in biology. For example, spectral graphs have been studied extensively in biological networks [31, 32].

The list of metrics in Table 8.1 hopefully provides a conceptual introduction to uninitiated readers who are not familiar with the nuances of complex networks. It is not meant to be an exhaustive list. It must be mentioned that there are a number of other network metrics like closeness centrality [12], eigenvector centrality [33], subgraph centrality [34], bipartivity [35], information centrality [12] etc.

**Table 8.1** Easy summary of common network metrics and concepts associated with each

Metric	Notion
Degree / connectivity	Number of connections a node has to other nodes in the network and perhaps also to itself
Geodesic distance	Length of the shortest path between a pair of nodes
Betweenness centrality	Fraction of all shortest paths which pass through a node; captures flow in information networks
Closeness centrality	Inverse of sum of distance of a node to all other nodes; denotes “closeness” to other nodes
Clustering coefficient	High for nodes in real-world networks; ratio of number of triangles connected to vertex to number of triples centered on vertex
Degree assortativity	Quantifies likeness of connections via pearson correlation coefficient, e.g. whether high-degree nodes or hubs are predominantly connected to other high-degree nodes or hubs in the network or to low-degree nodes
Degree gini coefficient	Similar to assortativity. Defined as $G(k) = \frac{\sum_{i=1}^N \sum_{j=1}^N  k_i - k_j }{N^2(k)}$ ; useful if there are outliers in degree distribution
k-core	Subgraph constructed by iteratively pruning all vertices of the network with degree less than $k$
Rich-club coefficient	Ratio, for every degree $k$ , of the number of actual to the number of potential edges for nodes with degree greater than $k$ , where $N_k$ is the number of nodes with degree larger than $k$ , and $E_k$ be the number of edges among those nodes. $\phi(k) = 2E_k/N_k(N_k - 1)$

### 8.3 Multiple Network Metrics

Now that we have a notional foundation about network metrics, it is time to discuss the relative importance of these network metrics. Degree, hubs and scale-free networks are already over represented in network literature [36]. However, while degree is certainly important in some circumstances, they are not always the only important metric. For example Fig. 8.1 clearly establishes that betweenness is more important than degree, in a number of scenarios. Again, assortativity is important in some circumstances [26, 27] and other metrics might play a significant role in other situations. Thus a question which naturally arises as to how we can identify which metrics are important in a given scenario and which ones are redundant.

In recent literature, this issue has been adequately addressed by the introduction of an appropriate quantitative framework [37, 38]. These papers demonstrated how and why multiple metrics and higher moments of some of these should be simultaneously studied in complex networks; they consider a significant number of network metrics, including higher moments of metrics, wherever appropriate. The first few moments of many distributions often (albeit not always), quantify a distribution sufficiently. Therefore, distributions of metrics like geodesic, betweenness, degree or clustering should be studied in depth whenever possible because they might carry important information about the system. Data mining techniques such as statistical dimension

reduction techniques like Principal Component Analysis (PCA) [39] and clustering can then be used for the identification of informative and redundant network metrics.

These papers clearly demonstrate that it is not just the degree or betweenness or some other metric which is important in every scenario. Most of the meaningful information is actually carried by a linear combination of some metrics and/or the higher moments of a few metrics [37, 38].

The power of these methods is demonstrated by the fact that they can also be used for comparing various network growth models among themselves and to detect how individual models compare with respect to real word data [37]. At this point, one might question if the consideration of the first few moments of a distribution is more an academic than a practical exercise. We will hopefully have an answer to this question by the end of this section.

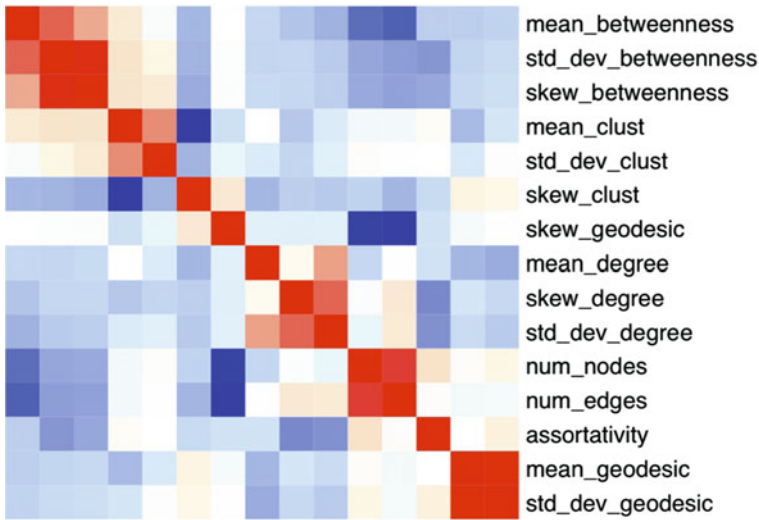
To demonstrate this we proposed a new model of network growth which was called the “graphlet model of network growth”. The motivation for proposing this model came from the observations that there are several instances in nature and science, where network growth via “graphlets” is observed. For example, in biology, gene duplication can add subnetworks to the network in the evolution of biological networks [40]. Again, in developmental transcriptional gene regulation, a mutation in a master regulator can add or eliminate whole pathways [41]. Computer software networks (composed of interacting functions or classes) often grow by the simultaneous addition of small groups of related elements. For example, (1) good design principles call for classes to be added in small groups called design, and, (2) to allocate, use, and free a resource (such as a file) functions are usually added together patterns in object-oriented languages [42]. However, most network growth models do not emphasize on *connected node arrivals* but on growth models like the Barabasi-Albert model [43]; which have one or more than one nodes arriving at every instant of time and attaching to the network by some defined mechanism.

It is apparent that this graphlet model will always produce trees and will not be able to capture the properties of many real world networks. Hence a new model was introduced where the incoming graphlet would throw  $l$  edges at random with probability  $\beta$  at each time step.

We then introduced a 15-dimensional attribute vector of seven well-known network properties. It is obvious that being armed with such a suite of metrics, would enable a very comprehensive and general comparison between any set of networks. These properties were: the number of nodes, the number of edges, the geodesic distribution, the betweenness coefficient distribution, the clustering coefficient distribution, the assortativity, and the degree distribution of the network. For the four distributions, the mean, standard deviation, and skewness were used as proxy attributes, adding up to a total of 15 attributes. Normalizing each value by subtracting the attribute mean and dividing by the attribute’s standard deviation, networks are mapped to points in a 15-dimensional space defined by these attributes (Fig. 8.2).

Heatmaps are extensively used across the sciences. They are typical tools for clustering data. Due to the hierarchical clustering used in a heatmap, the rows and columns get so ordered that the most correlated metrics are placed closest to each other. Clusters of *similar* network attributes can be identified by detecting blocks





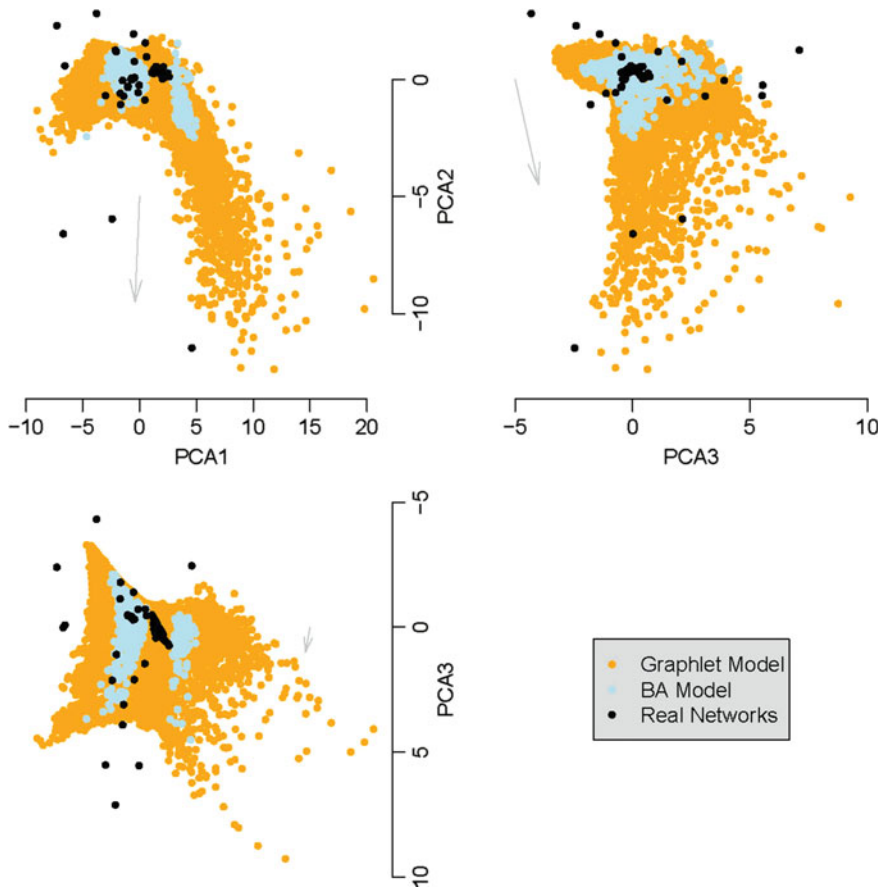
**Fig. 8.2** Symmetric heatmap of attribute correlations among networks. *Red (blue)* indicates perfect correlation (anti-correlation). *White* is the intermediate case of no correlation. The small amount of clustering along the diagonal attests to the relative independence of the attributes. *Source* [37]. Reprinted with permission from European Physical Society

of squares along the diagonal of the heatmap. Sizable blocks along the diagonal would denote redundancy. At the other end, a limited amount of clustering along the diagonal would imply that most of the network metrics we have chosen are effectively independent. Hence they would be informative for our analysis.

For our work, we assembled a collection of 113 diverse real-world networks from biological, social, and technical domains. This collection includes software call graphs, a social network of software developers, political social networks, three gene networks, three protein-protein interaction networks, cellular networks for several organisms, and several others downloaded from a web repository of networks.

To conduct a comprehensive comparison across many network metrics, we use a well known statistical dimension-reduction technique, like Principal Components Analysis [39]. The PCA algorithm ensures that when high dimensional data is projected to a lower dimension, the maximum variance of the dataset is retained. PCA finds the projection of an  $n$ -dimensional dataset onto an equidimensional space, such that the “new axes” (in other words, the principal components) are orthogonal and linear combinations of the original dimensional variables. The whole exercise is such that the first  $d$  axes, where  $d \leq n$  retain the maximal variance of the original data set possible with that many dimensions.

Figure 8.3 shows a projection of our model networks (orange points), the Barabasi-Albert model networks (light blue), and real-world networks (black) onto the first three principal components of the eleven-dimensional PCA space of our real-world data set. We omitted the number of nodes, skew of the degree distribution, and



**Fig. 8.3** A projection of our model networks (*orange points*), the Barabasi-Albert model networks (*light blue*), and real-world networks (*black*) onto the first three principal components of the eleven-dimensional PCA space of our real-world data set (we omitted the number of nodes, skew of the degree distribution, and variance and skew of the betweenness distribution from the original 15 attributes). Here, the PCA1 axis is primarily composed of (in terms of their coefficient's magnitude) a combination of the number of edges, mean and skew of geodesic, mean and standard deviation of clustering, and mean and standard deviation of degree. PCA2 is mainly a combination of the standard deviation and mean of geodesic, and assortativity. PCA3 is mainly a combination of the mean of betweenness, mean of geodesic, number of edges, and standard deviation of degree. As an example of a spread along an original parameter, the *gray arrow* is parallel to and shows the direction and magnitude of assortativity when projected onto this space. *Source* [37]. Reprinted with permission from European Physical Society

variance and skew of the betweenness distribution from the original 15 attributes for reasons discussed in [37]. Here, the PCA1 axis is primarily composed of (in terms of their coefficient's magnitude) a combination of the number of edges, mean and skew of geodesic, mean and standard deviation of clustering, and mean and standard deviation of degree. PCA2 is mainly a combination of the standard deviation and

mean of geodesic, and assortativity. PCA3 is mainly a combination of the mean of betweenness, mean of clustering, number of edges, and standard deviation of degree. As an example of a spread along an original parameter, the gray arrow is parallel to and shows the direction and magnitude of assortativity when projected onto this space. These principal components retain 71 % of the original data variance and demonstrate the larger coverage potential of the extended graphlet arrival model.

We end this section with the hope that we have been able to answer with clarity the question as to why we should be considering the higher moments of metric distributions. That they are indeed informative is reflected by the emphatic presence of a number of higher moments of metric distributions in the first few principal components [37]. Subsequent works in literature using similar approaches and techniques have also arrived at similar conclusions [44–46].

## 8.4 From Network Metrics to Organism Phenotypes

We then build on the methods developed in Sect. 8.3. The question we would like to ask is whether phenotypes and other biological properties of organisms depend crucially on the topological properties of their networks. And somewhat more ambitiously, does network topology encode for biological phenotype?

Towards this end, we used metabolic networks of 32 different microbes [38] based on data deposited in the What Is There (WIT) database [47]. This database contains metabolic pathways that were predicted using the sequenced genomes of several organisms. In these networks, edges represent sequences of reactions in the organisms cells while the nodes are enzymes, substrates, and intermediate complexes. The following three microbial species: *Actinobacillus actinomycetemcomitans*, *Rhodobacter capsulatus*, and *Methanobacterium thermoautotrophicum* were excluded from the original collection. This was because many of the phenotypic data or microbe characteristics do not seem to be publicly available for them. The network sizes vary from 2,982 nodes and 7,300 edges to 595 nodes and 1,354 edges.

The microbe characteristics or phenotypes that were investigated in our work are (1) microbe class (MC), (2) genome size (GS), (3) GC content (GC), (4) modularity (Q), (5) number of such modules ( $N_Q$ ), (6) motility (MO), (7) competence (CO), and whether these microbes are (8) animal pathogens (AP), (9) strict anaerobes (AN), or (10) extremophiles (EX). As is well-known, microbes are classified as bacteria or archaea. Genome size refers to the sum total of DNA contained within one copy of a genome. It is usually measured in the total number of nucleotide base pairs (commonly as millions of base pairs or mega-bases) or in terms of mass in picograms. Few microbes have much more DNA compared to other microbes; thus, an organism's genome size is not directly proportional to its complexity.

The percentage of nitrogenous bases on a DNA molecule, which is either cytosine or guanine and not thymine or adenine gives us the GC content. The data for GC content and genome size were obtained from the National Center for Biotechnology Information (NCBI) Entrez genome project database. Modularity of a biological net-

**Table 8.2** Exploring the association of microbe characteristics and phenotypes with network metrics

	Range	$\rho_{best}$	$\langle \rho_{rand} \rangle$	p-value	Best model variables
MC	Binary	0.113	0.507	$<3 \times 10^{-5}$	$N, E, geo_1, geo_2, geo_3, bet_1, bet_2, bet_3, deg_1$
GS	(0.58, 6.3)	0.476	1.302	$<10^{-6}$	$N, E, betw_1, bet_2, bet_3, deg_2, deg_3$
GC	(28.2, 66.6)	0.763	1.158	$<9.8 \times 10^{-5}$	$N, E, geo_1, geo_2, geo_3, bet_1$
Q	(0.59, 0.69)	0.005	0.033	$<10^{-6}$	$N, E, geo_2, geo_3, bet_1, bet_3, deg_1, deg_2$
$N_Q$	(14, 35)	2.102	6.413	$<10^{-6}$	$N, E, geo_1, geo_2, geo_3, bet_1, deg_1, deg_2$
MO	Binary	0.315	0.577	$<1.4 \times 10^{-5}$	$N, E, betw_3, deg_1, deg_2, deg_3$
CO	Binary	0.158	0.683	$<9 \times 10^{-6}$	$N, E, geo_1, geo_2, geo_3, betw_1, bet_3, deg_1, deg_2, deg_3$
AP	Binary	0.325	0.567	$<10^{-6}$	$geo_1, geo_2, betw_3, deg_2, deg_3$
AN	Binary	0.359	0.495	$<2.66 \times 10^{-4}$	$E, geo_1, geo_3, bet_1, bet_2, bet_3, deg_3$
EX	Binary	0.284	0.540	$<10^{-6}$	$geo_1, geo_2, bet_3, deg_1, deg_2, deg_3$

Microbe class (MC); Genome size (GS); GC content (GC); Modularity ( $Q$ ); Number of modules ( $N_Q$ ); Motility (MO); Competence (CO); and whether the microbes are Animal pathogens (AP), Strict anaerobes (AN) or Extremophiles (EX).  $N, E$  denote the number of nodes and edges in the network while  $geo_i, bet_i, deg_i$  denote the  $i$ 'th standardized moment of the network geodesic, betweenness and degree distributions respectively. The range of a values of a phenotype is given in case it is not binary. *Source* [38]. Reprinted with permission from the American Physical Society

work is defined as the fraction of edges within modules less the expected fraction of such edges. A state-of-the art algorithm [48] was used to determine the community structure in networks. This algorithm incorporated the edge directionality. Prior network community structure algorithms ignored edge directionality and applied methods developed for community structure in undirected networks. Obviously a lot of valuable information contained in the edge directions is lost as a result of this. It is well-known that modularity has an intimate connection to function in Biology. The concepts of structural and functional modularity are well-defined and modules typically correspond to gene circuits or pathways. Motility allows microbes to move away from undesirable environs towards desirable ones. The ability of a cell to take up extracellular DNA from its environment is measured by its Competence. Those microbes that do not require oxygen for growth and may even die in its presence are called Anaerobic organisms. Organisms which require extreme physical or geochemical conditions, in which majority of life on earth cannot survive are known as Extremophiles. While GS, GC, Q,  $N_Q$  can take values within the range mentioned in Table 8.2, the rest of the microbe characteristics or phenotypes are binary e.g., a microbe is either aerobic or anaerobic; either archaea or bacteria and so on.

We use a suite of 11 complex network metrics, so as to comprehensively compare all 32 networks simultaneously. To start with by assuming an initial dependence of the organism phenotype or characteristics on all network metrics because we do not know a priori which ones associate better than other metrics.

We then iteratively proceed to prune variables whose absence improves or does not significantly alter the quality of the resulting model. This is done by minimizing the well-known Akaike information criterion,

$$\alpha = 2k - 2 \ln L \quad (8.6)$$

where,  $k$  is the number of parameters in the statistical model and  $L$  is the maximum logarithmic likelihood for the estimated model.  $\alpha$  is a standard measure in statistics allowing for selection among various nested models. It penalizes models having many parameters and scores a model based on its goodness of fit to the data. In this way, we arrive at our “basis set” containing the smallest number of independent, indispensable network metrics that can be linked with an organism phenotype. We then use the root-mean-square error  $\rho$  to measure the goodness of fit of our model and the experimental data.  $\rho$  of an estimator  $\hat{Y}$  with respect to the estimated parameter  $Y$  is defined as the square root of the mean squared error,

$$\rho(\hat{Y}) = \sqrt{E[(\hat{Y} - Y)^2]} \quad (8.7)$$

We also report the significance of the best model, which we discussed above by bootstrapping with respect to the same model and using a random permutation of the observed data.

Table 8.2 enlists the complete results. We enumerate  $\rho$  of these random models,  $\rho_{rand}$ , and how many times (or whether at all)  $\rho_{rand} < \rho_{best}$ .  $\rho_{best}$  is obviously  $\rho$  of the best model. The normalized significance reflects the number of times  $\rho_{rand} < \rho_{best}$ . We observe  $10^6$  such random permutations, for each microbe phenotype. We also performed an analysis of variance of the difference of the model with all 11 variables and our model with fewest dependent variables. The difference is not significant.

For half of the microbe phenotypes in this study, namely, GS, Q, NQ, AP, and EX, we do not come across a single instance where  $\rho_{rand} < \rho_{best}$  for that phenotype. For each of these five phenotypes and also for the rest of the ones considered in this study,  $\rho_{rand} < \rho_{best}$ , with very low  $p$ -values.

Thus we can say with a good amount of confidence that there exists a strong association of organism phenotypes with relevant topological network metrics. As observable readily from Table 8.2, the presence of more topological network metrics does not necessarily enhance the prediction quality.

As mentioned before, an organism’s genome size is not directly proportional to its complexity. Therefore, it is interesting to observe in Table 8.2 that the association between topological metrics of the networks and their genome sizes is among the strongest of all phenotypes explored in this work.

## 8.5 Lessons Learnt

Networks are effective maps of complex systems. Traditionally biologists have focussed highly on degree in networks. In the process, they have largely ignored other metrics like betweenness, closeness etc which carry a lot of information about the system. The importance of these networks have however been known to sociologists for decades [12]. It would however, be a gross generalization to say that studies on biological networks have so far completely neglected other metrics. Nev-

ertheless, the suggestive examples presented at the start of this chapter followed with the exhaustive analysis of empirical data buttress our argument that it is only fair and wise; not to neglect the effect of other network metrics. A holistic picture of the system at hand can only be obtained via an exhaustive study of as many network metrics as possible.

## 8.6 Conclusion

We began this chapter with a discussion about how network metrics are important in systems biology and how topological analysis has often yielded novel and valuable insight in metabolic networks. New parameters like synthetic accessibility have demonstrated sufficient promise in predicting the viability of knockout strains with accuracy comparable to approaches using biochemical parameters (like FBA etc.) on large, unbiased mutant data sets [9]. We then discussed about the most common metric-degree. Albeit, degree is a very potent metric, we highlighted using simple examples the incompleteness of using only degree in network analysis. We next tried to conceptualise a commonly used centrality metric-betweenness. We then introduced assortativity or degree correlations which inform us as to whether the predominant pattern of connections in a network (hubs to hubs, hub to leaves or leaves to leaves). We emphasized on the fact that the list of metrics dealt with here is not exhaustive. However, the central point of this chapter is not to measure or enlist many network metrics. Rather, it is to emphasize the importance of the role of *simultaneous* use of *multiple* metrics. Using a suite of standard network metrics and armed with the higher moments of these metrics, we used standard tools from machine learning like clustering and principal component analysis to show that the information obtained from considering multiple network metrics allows us to make an in-depth comparison of networks and network growth models.

In conclusion, this chapter hopefully establishes that since networks play an important role in Systems Biology, it is only proper that we venture beyond traditional approaches of measuring just one or two hand-picked metrics. In fact, we expect that the use of proper quantitative techniques discussed in Sect. 8.3 in ways illustrated in Sect. 8.4, will lead to valuable insights not just for biological but for all types of complex networks.

## References

1. Albert R, Barabasi A-L (2002) Statistical mechanics of complex networks. *Rev Mod Phys* 74:47–97
2. Newman MEJ (2010) *Networks: an introduction*. Oxford University Press, Oxford, UK
3. Albert R, Jeong H, Barabasi A-L (2000) Error and attack tolerance of complex networks. *Nature* 406:378–382

4. Jeong H, Mason SP, Barabasi A-L, Oltvai ZN (2001) Lethality and centrality in protein networks. *Nature* 411:41–42
5. Varma A, Palsson BO (1994) Metabolic flux balancing basic concepts, scientific and practical use. *Biotechnology* 12:994–998
6. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 97:55285533
7. Segre D, Vitkup D, Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA* 99:1511215117
8. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420:190–193
9. Wunderlich Z, Mirny LA (2006) Using the topology of metabolic networks to predict viability of mutant strains. *Biophys J* 91:2304–2311
10. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51:661–703
11. Doyle JC et al (2005) The robust yet fragile nature of the Internet. *Proc Natl Acad Sci USA* 102:14497–14502
12. Wasserman S, Faust K (1994) *Social network analysis*. Cambridge University Press, Cambridge, UK
13. Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40:35–41
14. Guimera R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. *Proc Natl Acad Sci USA* 102:7794–7799
15. Wuellner DR, Roy S, D'Souza RM (2010) Resilience and rewiring of the passenger airline networks in the United States. *Phys Rev E* 82:056101
16. Dunn R, Dudbridge F, Sanderson CM (2005) The Use of Edge-Betweenness Clustering to Investigate Biological Function in Protein Interaction Networks. *BMC Bioinform* 6:39
17. Hahn MW, Kern AD (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol* 22:803–806
18. Hegde SR, Manimaran P, Mande SC (2008) Dynamic changes in protein functional linkage networks revealed by integration with gene expression data. *PLoS Comput Biol* 4(11):e1000237
19. Joy MP, Brock A, Ingber DE (2005) Huang S high-betweenness proteins in the yeast protein interaction network. *J Biomed Biotech* 2:96–103
20. Liu J et al (2009) Analysis of drosophila segmentation network identifies a JNK pathway factor overexpressed in kidney cancer. *Science* 323:1218–1222
21. Liu YY, Slotine J-J, Barabasi A-L (2011) Controllability of complex networks. *Nature* 473:167–173
22. Banerjee SJ, Roy S (2012) Key to network controllability arxiv:1209.3737
23. Newman MEJ (2002) Assortative mixing in networks. *Phys Rev Lett* 89:208701
24. Reynolds-Feighan A (1998) The impact of U.S. airline deregulation on airport traffic patterns. *Geogr Anal* 30:234
25. Sen A (1973) *On economic inequality*. Clarendon Press, Oxford, UK
26. Bagler G, Sinha S (2007) Assortative mixing in protein contact networks and protein folding kinetics. *Bioinformatics* 23:1760–1767
27. Pechenick DA, Payne JL, Moore JH (2012) The influence of assortativity on the robustness of signal-integration logic in gene regulatory networks. *J Theo Biol* 296:21–32
28. Colizza V, Flammini A, Serrano MA, Vespignani A (2006) Detecting rich-club ordering in complex networks. *Nat Phys* 2:110115
29. Zhou S, Mondragon RJ (2004) The rich-club phenomenon in the internet topology. *IEEE Commun Lett* 8:180
30. van den Heuvel MP, Sporns O (2011) Rich-club organization of the human connectome. *J Neurosci* 31:15775–15786
31. Banerjee A, Jost J (2009) Graph spectra as a systematic tool in computational biology. *Discrete Appl Math* 157:2425–2431



32. Perkins AD, Langston MA (2009) Threshold selection in gene co-expression networks using spectral graph theory techniques. *BMC Bioinform* 10(Suppl 11):S4
33. Newman MEJ (2008) Mathematics of networks. In: Blume LE, Durlauf SN (ed) *The new palgrave encyclopedia of economics*, 2<sup>nd</sup> edn. Palgrave Macmillan, Basingstoke
34. Costa LDF, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: a survey of measurements. *Adv Phys* 56:167242
35. Estrada E, Rodriguez-Velazquez JA (2005) Spectral measures of bipartivity in complex networks. *Phys Rev E* 72:046105
36. Roy S (2012) Systems biology beyond degree, hubs and scale-free networks. *Syst Synth Biol* 6:31–34. doi:[10.1007/s11693-012-9094-y](https://doi.org/10.1007/s11693-012-9094-y)
37. Filkov V, Saul ZM, Roy S, D'Souza RM, Devanbu PT (2009) Modeling and verifying a broad array of network properties. *EPL (Europhys Lett)* 86:28003
38. Roy S, Filkov V (2009), Strong associations between microbe phenotypes and their network architecture. *Phys Rev E* 80: 040902 (R).
39. Jolliffe IT (2002) *Principal component analysis*, 2<sup>nd</sup> edition. Springer-Verlag, New York
40. Kashtan N, Alon U (2005) *Proc Natl Acad Sci USA* 102:13773
41. Davidson EH (2006) *The regulatory genome: gene regulatory networks in development and evolution*. Academic Press, Elsevier, San Diego
42. Gamma E et al (1995) *Design patterns: elements of reusable object-oriented software*. Addison-Wesley Longman Publishing Co., Boston
43. Barabasi A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509
44. Agarwal S, Villar G, Jones NS (2010) High throughput network analysis. In: *Proceedings of the workshop on analysis of complex networks (ACNE), European conference on machine learning and principles and practise of knowledge discovery in databases (ECML, PKDD)*. Barcelona, Spain, pp 13–18
45. Bounova G, de Weck O (2012) Overview of metrics and their correlation patterns for multiple-metric topology analysis on heterogeneous graph ensembles. *Phys Rev E* 85:016117
46. Villar G, Agarwal S, Jones NS (2010) High throughput network analysis in machine learning in systems biology (MLSB). In: *Proceedings of the 4th international workshop*. Edinburgh, Scotland, pp 5–7
47. Overbeek et al (2000) *Nucl Acids Res* 28:123
48. Leicht EA, Newman MEJ (2008) *Phys Rev Lett* 100:118703



# Chapter 9

## Understanding and Predicting Biological Networks Using Linear System Identification

Alberto Carignano, Ye Yuan, Neil Dalchau,  
Alex A. R. Webb and Jorge Gonçalves

**Abstract** This chapter demonstrates how linear systems can be used to model biochemical networks. Such models give predictable power that can be used to generate hypotheses, which in turn can be (in)validated experimentally. The advantages of linear systems are that they are relatively simple, efficient to obtain and simulate, and have been studied in great detail. In spite of inherent nonlinearities in real world applications, linear systems have been successfully used in control theory as a tool to model, analyse and control technological systems. In contrast, although at the molecular level reactions are nonlinear, modelling of key behaviours important to predict new features of a system can in many instances be captured by linear dynamics. Guided by a simple example, this chapter explains step-by-step how to use linear system identification (SID) to obtain causal relationships between different biological species in complex networks. We will cover key aspects of model

---

A. Carignano · Y. Yuan · J. Gonçalves (✉)  
Control Group, Department of Engineering, University of Cambridge, Cambridge, UK  
e-mail: jmg77@cam.ac.uk

J. Gonçalves  
University of Luxembourg, LCSB, Luxembourg  
e-mail: jorge.goncalves@uni.lu

A. Carignano  
e-mail: ac737@cam.ac.uk

Y. Yuan  
e-mail: yy311@cam.ac.uk

A. A. R. Webb  
Department of Plant Sciences, University of Cambridge, Cambridge, UK  
e-mail: aarw2@cam.ac.uk

N. Dalchau  
Biological Computation Group, Microsoft Research,  
Cambridge, UK  
e-mail: ndalchau@microsoft.com

estimation, validation and selection. The corresponding Matlab™ codes will be also be introduced. The chapter ends with illustrations of practical applications through two case studies, where SId has been used to further our understanding of biological networks.

**Keywords** Linear systems · Nonlinearity · System identification · Linear system identification · Molecular · Biomolecular · Model selection · State · System · Input · Cholesterol

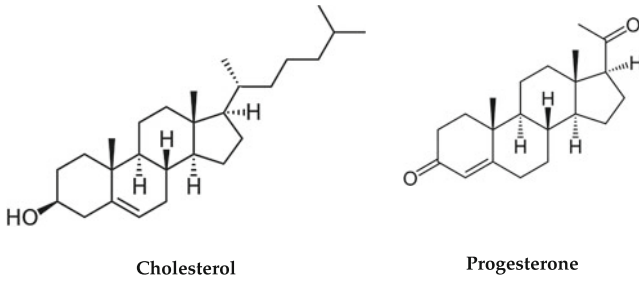
## 9.1 Introduction

The use of computational and algorithmic approaches in biological research has greatly increased in the last decade. The development of new technologies is creating increasingly larger datasets that describe biological responses to genetic or pharmacological perturbation. Consequently, there exists a pressing need to process and interpret these data, despite often having only rudimentary knowledge of the interactions between the constituent components involved in the biological system of interest.

A spectrum of methodologies and models has arisen to help tackle this challenge, from purely statistical and correlative approaches to data processing and hypothesis testing, through to sophisticated agent-based simulations at the molecular or cellular level. Models also range between static (steady-state) descriptions through to dynamical models, deterministic or stochastic, and spatially homogeneous or heterogeneous. The level of abstraction that is chosen is normally a consequence of the level of prior knowledge one has of the system of interest, and the question being addressed.

System identification originates in control theory, and is aimed at characterising specific model classes (systems that include only linear terms, for example) from observational data alone [7]. In particular, the response of a *system* to an *input*, can be described with such methods. This approach enables efficient inference of the model parameters, which quantify the interactions between model *states* (system components). Applied to a biomolecular system, this approach offers a way of deriving a dynamical model in which the interactions between the constituent molecules are unspecified. From such a model, predictions can be made about the dynamical behaviour in alternative conditions (different input signals), offering the means to investigate a previously uncharacterised biological network, and generate testable hypotheses about the functioning and organisation of the system.

In this book chapter, we demonstrate how system identification can be applied to linear models of biological systems. A tutorial introduces the use of the Matlab System Identification toolbox to achieve model estimation and cross-validation [7] through the example of hormone biosynthesis. Following this, two case studies are introduced in which system identification has led to demonstrable insights into the



**Fig. 9.1** Chemical representation of cholesterol and progesterone molecules

functioning of partially characterised biological mechanisms: the NF- $\kappa$ B activation network [3] and the regulation of cytosolic-free calcium ( $[Ca^{2+}]_{\text{cyt}}$ ) by the circadian clock in *Arabidopsis* [4].

## 9.2 A Tutorial on System Identification: Linear Modelling of Progesterone Biosynthesis

### 9.2.1 Background

We start by considering a well characterised biological system, the production of progesterone from cholesterol. We illustrate how linear SID can be used to study such a reaction. Cholesterol is a fat molecule that can be ingested and synthesised by the cell, while progesterone is a fundamental female hormone involved in supporting gestation [9] (Fig. 9.1)

The conversion of cholesterol into progesterone is a single step of the steroidogenesis metabolic pathway, which generates steroids from cholesterol. All the elements of this pathway have been studied extensively and a complete dynamical description of the reaction can be obtained from the literature [9]. Therefore, this pathway offers an ideal case study for testing novel modelling methodologies. We assume no prior knowledge and try to deduce a model from time-series experimental measurements alone.

A natural question that arises when considering the causality of the reaction is ‘how strongly does cholesterol influence the rate of progesterone synthesis?’ In particular, what are the forward and reverse reaction rates of progesterone–cholesterol interconversion? In chemical reaction notation, we write



where  $X_1$  and  $X_2$  represent cholesterol and progesterone respectively.

Assuming that the reaction (9.1) follows mass-action kinetics, then we can write down mathematical equations that define how the abundance of progesterone and cholesterol are related in terms of the forward and reverse reaction rates  $k_1$  and  $k_2$ . Let  $x_1$  and  $x_2$  represent the concentration of cholesterol and progesterone, respectively. The reaction formula (9.1) tells us that  $x_1$  increases at a rate  $k_2x_2$  and decreases at rate  $k_1x_1$ , while the opposite can be said for  $x_2$ . Therefore, the rate of change of  $x_1$  with respect to time,  $\frac{dx_1}{dt}$ , is equal to  $k_2x_2 - k_1x_1$ . Hence reaction (9.1) can be written as a system of differential equations:

$$\frac{dx_1}{dt} = -k_1x_1 + k_2x_2 \quad (9.2)$$

$$\frac{dx_2}{dt} = k_1x_1 - k_2x_2 \quad (9.3)$$

Note that although data is collected at discrete times, systems are modelled in continuous-time. There are two fundamental reasons for this. First, biochemical systems evolve in continuous time. Second, modelling the system in discrete time can lead to very different parameter values, which in turn can lead to wrong conclusions. A discrete time model assumes that the system has been sampled fast enough to capture the whole dynamic. This assumption is often difficult to justify, especially in a biological process where it can be hard or impossible to isolate all the active components.

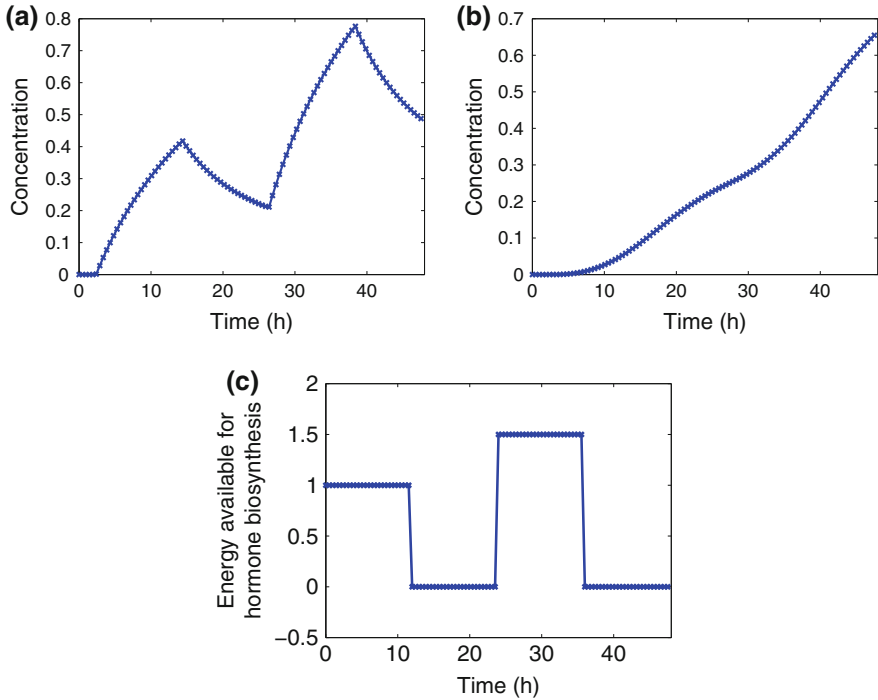
Control theory is concerned with how systems respond to external forcing. Biological systems also respond to external forcing, so we introduce the concept of a control variable. For simplicity, assume there is a linear relationship between the amount of food eaten and the amount of cholesterol produced, i.e. we can ‘control’ the amount of  $x_1$  by the amount of food we eat; we will call the energy derived from food consumption  $u$ . With the control variable, the equations become:

$$\frac{dx_1}{dt} = -k_1x_1 + k_2x_2 + k_3u \quad (9.4)$$

$$\frac{dx_2}{dt} = k_1x_1 - k_2x_2 \quad (9.5)$$

where  $k_3$  is the correlation between available energy for hormone biosynthesis and cholesterol increase. Equations (9.4), (9.5) represent the model under investigation and contain 3 parameters that need to be estimated.

For the sake of this exercise, we will only use simulated data to represent collected measurements of the concentrations of  $x_1$ ,  $x_2$  and the amount of food consumed during a 48 h period (measured as energy source for the reaction). The model used for the simulations follows reasonable biological assumptions and approximates the real biological system. For details on the model please see the Appendix. We are interested in estimating a model as close as possible to the original one using the simulated data only. The time-series data obtained from the simulation are displayed in Fig. 9.2.



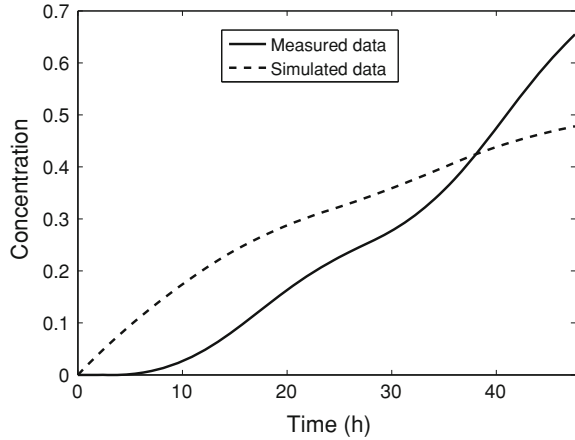
**Fig. 9.2** **a** Cholesterol concentration; **b** progesterone concentration; **c** energy available for the biochemical process. Data was collected every 5 min over 48 h

The question now is: how do we estimate the parameters in the linear model described by Eqs. 9.4 and 9.5? A simple solution is to use literature and biological intuition to try and guess ‘reasonable’ parameters, simulate the system and compare with the data. For example, using the parameter values  $k_1 = 1$ ,  $k_2 = 2$  and  $k_3 = 0.5$  and initial condition  $(1, 0)$ , we obtain an increase in the concentration of progesterone over time (Fig. 9.3).

It is clear that these parameter values do not capture all dynamics in the real data. While both concentrations of progesterone increase over time, there is a clear mismatch between the simulation and the data. Mathematically, there are many ways to quantify and to make this mismatch precise. A standard approach is the least squares error, which is the sum of the squared differences between each experimental datapoint and corresponding simulation. If there are  $N$  datapoints, then the least squares error can be written as

$$V_N = \sum_{k=1}^N (y_k - \hat{y}_k)^2 \tag{9.6}$$

**Fig. 9.3** Simulation of the linear model for progesterone biosynthesis, with the system parameter values set to  $k_1 = 1$ ,  $k_2 = 2$  and  $k_3 = 0.5$  and initial conditions  $(1, 0)$ . The simulation (*dashed line*) is compared with the target system data (*solid line*)



where  $y_k$  represents the sequence of experimental datapoints and  $\hat{y}_k$  is the corresponding prediction sequence from a given model ( $\hat{y}_k$  is obtained by first simulating the continuous-time model and then sampling its response at the same times as the data; in this case, since the datapoints are evenly spaced,  $\hat{y}_k = \hat{y}(kT)$ , where  $T = 5$  s is the sampling interval).

To minimise the squared error defined by Eq. (9.6), we could change the parameters by hand until reaching a satisfactory solution. However, this method becomes prohibitively time-consuming for large systems. A more systematic method is to define an optimisation problem where we seek to minimise the squared error in Eq. (9.6) given a particular model class. In our example, the model class is that of two dimensional linear systems. To solve the optimisation problem, we can use the ‘prediction-error method’ (PEM), a technique widely used in Sid [7].

PEM usually starts from a random estimate of the parameters and initial conditions of the model (though these can optionally be initialised). For each datapoint, a prediction is formed from the previous datapoint(s), and the squared error is computed between the predicted data points and the target data. The algorithm then iterates towards parameters that reduce the squared error, stopping when no further improvement can be made. For more details about the method used in the minimization process, see [7]. The Matlab™ function ‘pem’ returns parameter estimates and initial conditions, by using the following code:

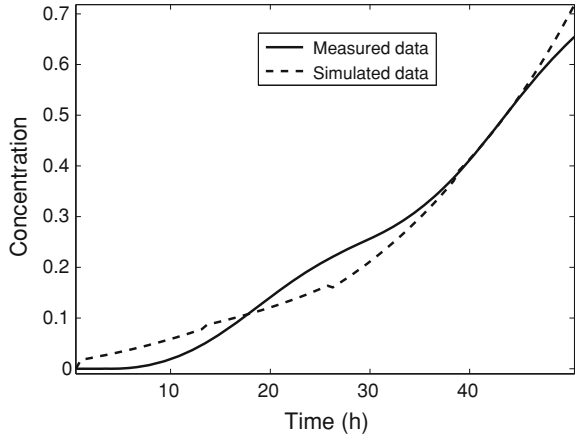
$$\text{data} = \text{iddata}(y, u); \quad (9.7)$$

$$m = \text{pem}(\text{data}, 2); \quad (9.8)$$

$$m = \text{dtc}(m); \quad (9.9)$$

where the function `iddata` inputs the data and the second entry of the `pem` function provides the estimated model order. By default, the function `pem` returns a discrete-time model, so we use the command `dtc` to transform the model from discrete to continuous-time.

**Fig. 9.4** Simulation of the estimated second order model for the progesterone concentration. Fitness = 84 %



After obtaining a model and simulating it, we can derive a definition of fitness from the cost function given by

$$\text{fitness} = 100 * \left( 1 - \frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{\sum_{k=1}^N (y_k - \bar{y})^2} \right) \tag{9.10}$$

where  $\bar{y}$  is the average value of the experimental data (in order to avoid divisions by zero, a different formula has to be used to estimate the fitness of a constant output). It is easy to check that a zero model error corresponds to a 100 % fit. The Matlab™ function ‘compare’ can be used to compute the fitness using:

$$[\sim, \text{fit}] = \text{compare}(\text{data}, \text{model}); \tag{9.11}$$

Simulation of the system using the estimated parameters produces the plot in Fig. 9.4.

By default, models estimated by ‘pem’ are in a ‘state-space’ form, which are matrix forms of Eqs. (9.4), (9.5). To clarify, consider the continuous-time linear system expressed by the equations:

$$\dot{x} = Ax + Bu \tag{9.12}$$

$$y = Cx + Du \tag{9.13}$$

where  $x$  are the ‘state variables’,  $u$  are the inputs to the system,  $A$  and  $B$  are matrices that contain the parameters corresponding to how the states affect each other and how inputs affect the states, respectively. The measurements are given by  $y = Cx + Du$ , which can be a linear function of the states (via matrix  $C$ ) and inputs (via matrix  $D$ ). In practice, the matrix  $D$  is usually 0. In our example,  $x_1$  and  $x_2$  are our states, food is the input  $u$ , and the matrices are given by

$$A = \begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix}, \quad B = \begin{bmatrix} k_3 \\ 0 \end{bmatrix}, \quad C = [0 \ 1], \quad D = 0$$

with this definition of  $C$ , the output is given by  $y = x_2$ .

Equations (9.12) and (9.13) describe a Linear Time-Invariant model (LTI): linear because  $Ax$  is a linear combination of the state variables ( $a_1x_1 + a_2x_2 + \dots + a_nx_n$ ), i.e. no mixed products  $x_i x_j$  or nonlinear functions, and time-invariant because the system parameters do not depend on time.

In general, the algorithm pem used in Matlab™ is given by

$$\begin{aligned} \text{data} &= \text{iddata}(y, u, Ts) \\ m &= \text{pem}(\text{data}, n) \end{aligned}$$

$Ts$  is the sample time of the two time series  $y$  and  $u$ , and  $n$  is the model order, i.e. dimension of  $A$ . If there is no input, simply write  $\text{data} = \text{iddata}(y, [], Ts)$ .

It is possible to impose additional structure on the model. For instance, in the cholesterol example if we were to measure  $x_2$  instead, then  $C = [0 \ 1]$ . We can set any entries in  $A$ ,  $B$ ,  $C$ ,  $D$ , either as fixed relations, or as initial guesses (e.g. if we have an idea about the value of an entry, then the optimisation process is more likely to converge faster).

For instance, let us assume that we have some evidence that the correct values for the matrices  $A$ ,  $B$ ,  $C$  and  $D$  are:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}; \quad B = \begin{bmatrix} 1 \\ 2 \end{bmatrix}; \quad C = [0 \ 1]; \quad D = 0; \quad (9.14)$$

We can then set these values as initial guesses for the optimisation problem using the command:

$$m = \text{idss}(A, B, C, D).$$

Suppose we would also like a specific structure (for instance  $C = [0 \ 1]$ ) to be preserved during the optimisation process. This is done by letting

$$m.Cs = [0 \ 1].$$

Equation (9.9) uses the function `dtc` to obtain a continuous-time system from a discrete-time system in (9.8). This can be done automatically by specifying that the sampling time is 0, i.e.  $m.Ts = 0$ . Finally, we call the function `pem` as

$$m1 = \text{pem}(\text{data}, m); \quad (9.15)$$

There is no need to specify the model order here because the model structure  $m$  is already implicitly setting it (from the dimension of the matrix  $A$ ).



## 9.2.2 Model Validation

System identification combines a number of techniques to solve model estimation problems. If a model is correctly identified, then it can be used to make predictions by simulating a wide range of conditions. This, in turn, can help improve the understanding of the system, and save time and money invested in experimental research.

In the previous section, we went through the estimation process and obtained a model with a high fit (84 %). This suggested that most of the dynamics of the system were captured. However, it is possible that the model is not describing the biochemical reaction properly and instead is spuriously reproducing the data with the wrong underlying model. This issue is called ‘over-fitting’ (for an example, [2, p. 7]) and it is one of the most common and relevant problems in mathematical modelling. A good model needs to be ‘flexible’ to correctly predict new experimental conditions. Such cross-validation tests can help to rule out spurious close fits to the data, and distinguish between good and bad models.

Suppose we collect new data using a different input (different amount of food in this case) and different initial conditions (different initial concentration of cholesterol and progesterone in the blood). Can the model correctly predict the dynamics of progesterone biosynthesis in the alternative conditions? A new dataset is in Fig. 9.5.

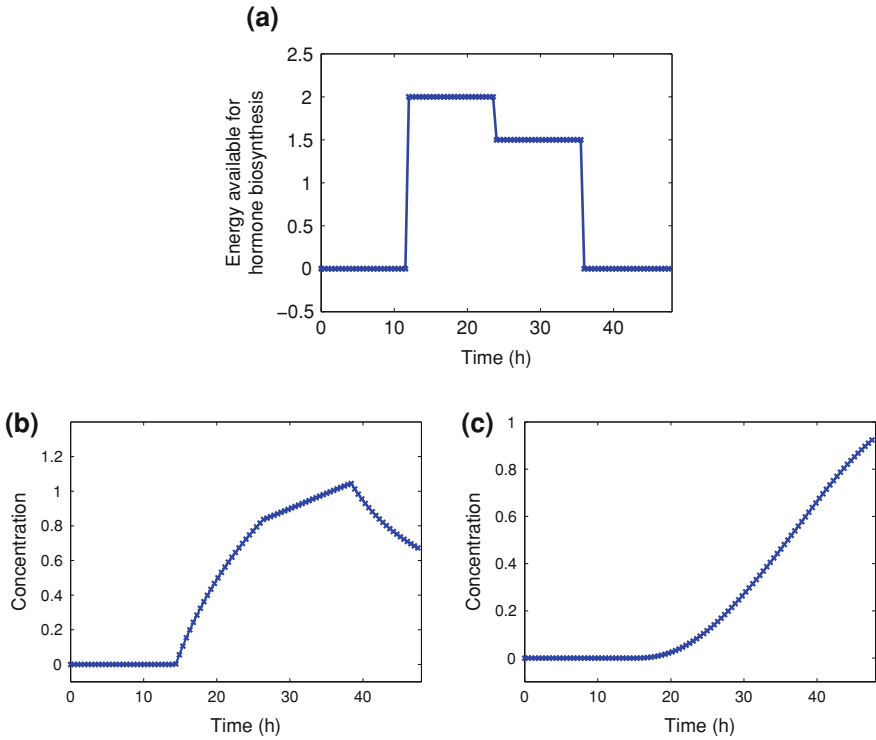
The simulation using this new input is shown in Fig. 9.6 [in Matlab™ this can be done using the command ‘`lsim(m,u,t)`’]. The model describes the data well, correlating well over the range of datasets.

What we have just done is called ‘Model Validation’ and it assesses whether a model has power or not. In this example, the model showed an acceptable predictive power. The next section returns to the question of over-fitting and will discuss trade-offs between precision and over-fitting. In particular, it will discuss the question of how to improve model prediction.

It is interesting to note that all the modelling we have done only used time-series data for progesterone, not cholesterol. This is one advantage of linear system identification: the model can be obtained from input-output data alone, with no need to measure every single state involved with the output (in our case we have three states—see Appendix). Once we have a model, only input data (i.e. exogenous quantities that excite the system in our hypothetical experiment) are needed for prediction (assuming zero initial conditions).

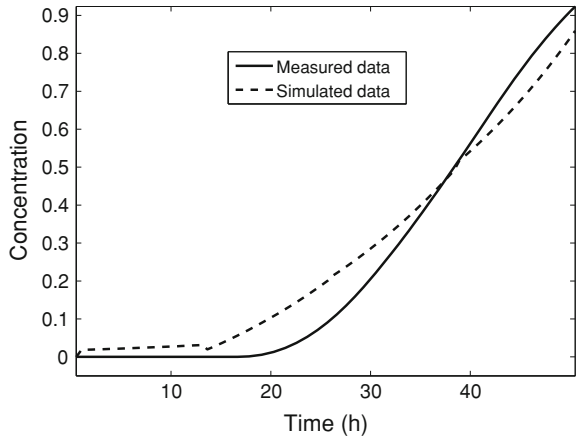
## 9.2.3 Hidden States and Model Selection

Consider again our starting example. The single reaction description (9.1) is a simplification of a more involved biochemical pathway. There is at least one intermediate state in the chain, which is the steroid hormone ‘pregnenolone’ (see Fig. 9.7). Incorporating pregnenolone into the chemical reaction model yields

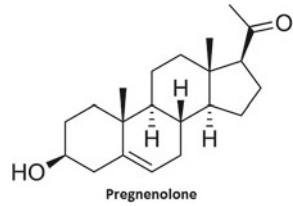


**Fig. 9.5** Validation dataset: **a** cholesterol concentration; **b** progesterone concentration; **c** food consumption. Data collected over 48 h

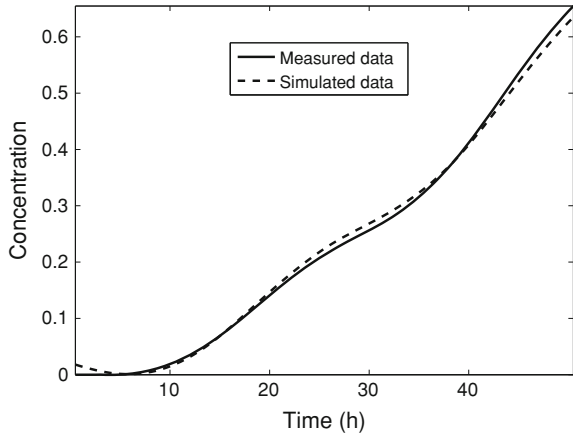
**Fig. 9.6** Validation of the estimated second order model for the new progesterone concentration dataset



**Fig. 9.7** Chemical representation of the pregnenolone molecule



**Fig. 9.8** Simulation of the estimated third order model for the progesterone concentration



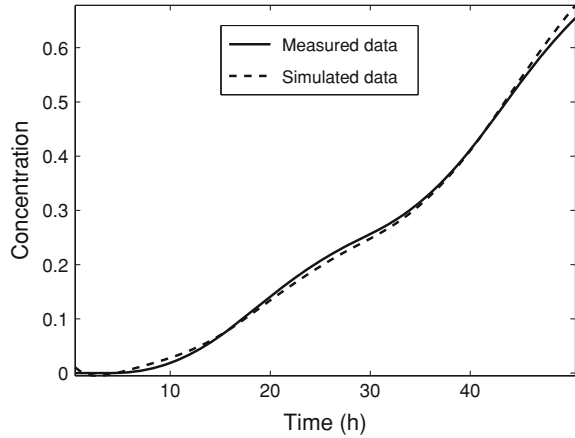
where  $X_1$  represents cholesterol,  $X_2$  pregnenolone and  $X_3$  progesterone.

Considering the action of pregnenolone, the actual model should be of at least 3 variables (3rd order). We can impose this structured condition on PEM with the command

$$m = pem(data, 3);$$

The new model generates the simulation shown in Fig. 9.8. It is clear that there is a much better fit between this model and the data than the single reaction 2nd order model. The new fit is 90 % (compared with 84 % for the 2nd order model). Since increasing the order improved the fitness, it seems natural to consider adding other intermediate steps in the chemical process: perhaps the real system is 4th order, with another hormone somewhere in the chain. If we estimate a 4th model for the data and simulate it, we obtain an even better fit (92 %; Fig. 9.9). It turns out that a 5th order model would have an even better performance (94 %), and so on. In fact, increasing the order of the model can only improve the overall fitness. A higher order system, however, implies more parameters to estimate, which in turn means more degrees of freedom. There are clear trade-offs here. How much information does the dataset

**Fig. 9.9** Simulation of the estimated fourth order model for the progesterone concentration



provide? How many parameters can we justifiably estimate? And when do we stop fitting the dynamics in the system and start fitting noise?

In the last example, we noticed how increasing the model order helped achieve a better fit of the data. These extra state variables are called ‘hidden states’ and represent intermediate steps in the biological process that we have no measurements of. The ability to infer the presence of intermediate reactions is a very powerful feature of linear SId. The more hidden states the model has, the more flexibility there is in the estimation process. However, if we are not careful, a model might end up with more states than the real system. This is why we need a method to select the order of the model. This process is called ‘Model Selection’ and it generally uses an information criterion that takes into account the goodness of fit and penalises the number of estimated parameters. Higher order models have naturally a better fit (higher flexibility) but also a bigger number of parameters.

A common approach to quantifying the compromise between goodness of fit and over-parametrisation are information criteria. The most common information criterion in the literature is the Akaike Information Criterion (AIC) [1], which is defined as

$$\text{AIC} = \log \bar{V}_N + \frac{2d}{N} + \frac{2(d - \log \bar{V}_N)}{N} \quad (9.17)$$

where  $\bar{V}_N$  is the value of the cost function [Eq. (9.6)],  $d$  is the number of estimated parameters and  $N$  is the number of datapoints in the set. Low values of AIC correspond to a good balance between small values of the cost function and small numbers of parameters. In practice, we compute the AIC for a range of different model orders and compare them. The AIC values should decrease every time a higher order model has a significant fitness improvement. When they stop decreasing, then we might have reached a good candidate. There will be a point past which additional parameters serve only to increase the fit to the data by a small amount, which translates

**Table 9.1** AIC coefficients corresponding to model of different orders

Order	AIC coefficient	Fitness (%)
2	0.13	84
3	-22.64	90
4	-22.55	92
5	-22.38	94

into an increase in the AIC score. This is an example of over-fitting, as discussed in Sect. 2.1.

Returning once more to the example, by fitting models of different orders we obtain the AIC coefficients in Table 9.1. This analysis suggests that the best model order is 3. Adding extra hidden states does not improve the AIC score. This is a consequence of not being able to significantly improve the fitness after the third order model.

### 9.2.4 System Identification for Noisy Measurements

So far we have seen how to estimate and validate a model, based on artificial data. In real life, however, noise is present at different levels in experimental data and can make modelling very challenging. Noise can be a consequence of inaccuracies in the measurement devices, stochastic process variations (intrinsic), environmental fluctuations (extrinsic), etc. In LTI systems, noise is typically incorporated into the model as:

$$\dot{x} = Ax + Bu + Ke \tag{9.18}$$

$$y = Cx + Dy + e \tag{9.19}$$

where the variable  $e$  is a random signal, typically assumed to be white (Gaussian) noise, i.e.  $e(t) \sim \mathcal{N}(0, \sigma^2)$ , where  $\sigma^2$  is the variance of the signal.

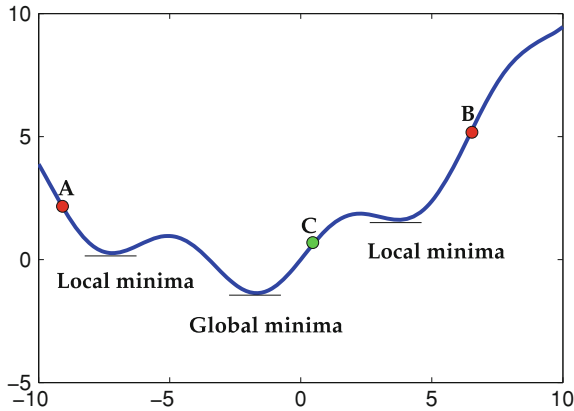
To take into account noise in the estimation of a model, we use the following code:

$$m = \text{idss}(A, B, C, D, K); \tag{9.20}$$

$$m2 = \text{pem}(\text{data}, n). \tag{9.21}$$

where  $K$  is a vector with the same number of columns as the number of states. In a sense, system identification seeks to distinguish between the true signal and noise. This feature is extremely relevant in a biological environment, where the observations are usually subject to different sources of noise.

In this more general setting, PEM is more likely to fail to converge. Noisy signals and high order models generate hard optimisation problems and cost functions with



**Fig. 9.10** The *blue line* represents the cost function; the subject of our minimization. The optimization algorithm works by taking a starting point on the line and then progressing ‘downhill’. If *A* or *B* are used as a starting point, the algorithm will terminate in the closest local minimum (wrong result). However using *C* as starting point will result in the correctly optimized solution

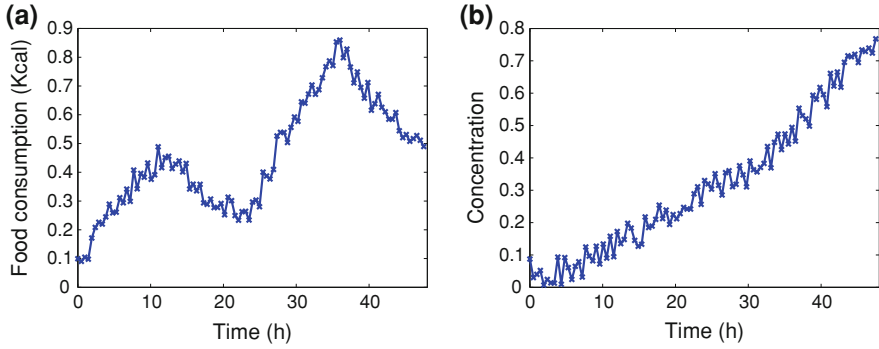
several local minima. A local minimum is a value that is the lowest in its neighbourhood, but not necessarily the lowest in the parameter space. Many optimisation algorithms methods work by moving ‘down hill’, and consequently are prone to getting stuck at local minima. Upon reaching a local minimum, the algorithm stops as it cannot ‘climb up’ back on top and look for a better minimum (see Fig. 9.10). The PEM method uses gradient-based optimisation, and so suffers from the drawbacks just described.

One way to get around local minima is to try several different initial conditions for PEM and then choose the estimated model with the highest fit (e.g. expectation maximisation (EM) algorithms [5]). Different starting points for the algorithm might ‘lead’ to paths that avoid local minima (see Fig. 9.10). While this still does not guarantee a global solution to the problem, it usually improves the performance with respect to a single PEM run.

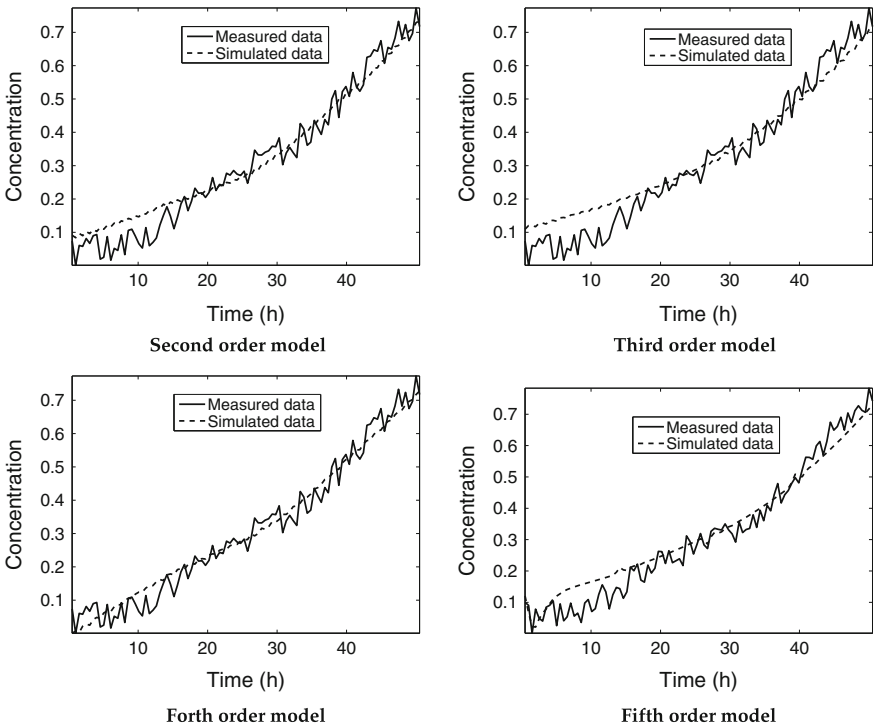
### 9.2.5 A Noisy Signal Example

Let us go back to our main example, this time with a more realistic scenario that includes noise (however, we use the same input data as before). Our time-series data could look like Fig. 9.11.

We define the model structure as in Eq. (9.20) and then use PEM to estimate the best model for orders from 2 to 5. Each model is estimated 20 times using different initial conditions to avoid local minima. Depending on the fit, the best candidates are selected for each order. Resulting simulations are shown in Fig. 9.12.

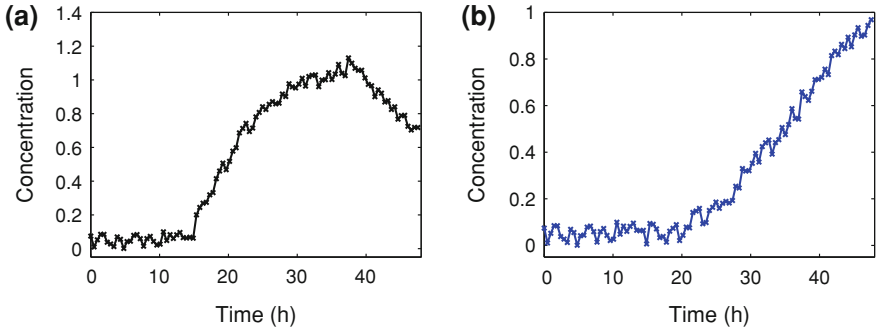


**Fig. 9.11** Noisy measurements: **a** cholesterol concentration; **b** progesterone concentration; data collected over 48 h

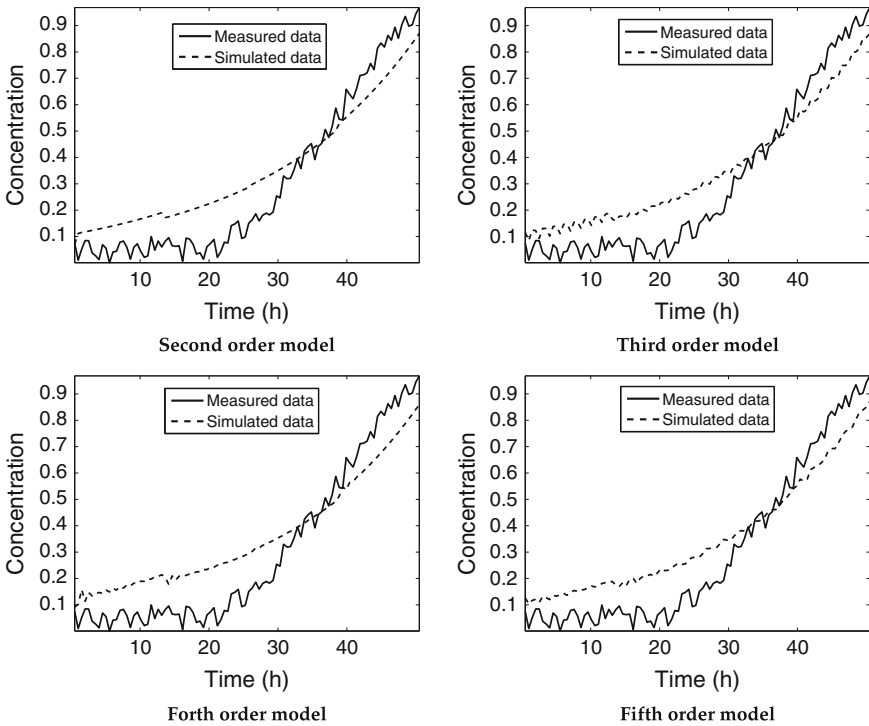


**Fig. 9.12** Simulations of progesterone concentration using noisy measurements and different model orders

To validate the models, we use a noisy version of our previous validation set (Fig. 9.13). The validation process in this case is really important as all the models are able to closely reproduce the estimation dataset. Simulations of the validation sets are depicted in Fig. 9.14.



**Fig. 9.13** Noisy measurements for validation data: **a** cholesterol concentration; **b** progesterone concentration; data collected over 48 h



**Fig. 9.14** Simulations of progesterone concentration for the validation dataset with different model orders

The process of model selection is not an easy task in this case. ‘By eye’, there does not appear to be a significant improvement between the models, suggesting we should favour simplicity and opt for the second-order model. Appealing to the quantitative metrics introduced above, we can use the ‘compare’ command in Matlab<sup>TM</sup>, and



**Table 9.2** Summary of the fitness for estimation and validation dataset and AIC coefficients (calculated for the estimation dataset) for model from the 2nd to the 5th order

Order	Fit for the estimation set	Fit for the validation set	AIC coefficient
2	83.16	62.46	-7.01
3	85.4	69.37	-7.028
4	85.67	69.47	-7.021
5	86.73	70.13	-6.99

calculate the AIC coefficient (see Table 9.2). There is only a small difference between the AIC coefficients of the 3rd and the 4th order models. Also, an extra state variable (5th order) does not add any extra information and the best AIC score is achieved using the 3rd order model. Therefore we are left with the choice of either 2nd order.

At this point, any further conclusions become subjective. One could argue that the delayed behaviour of progesterone) showed in the two data sets (both progesterone and cholesterol time series remain flat for the first 2 h despite a non zero input) is not correctly reproduced in a second order model. This feature is indeed more evident in the third order model simulation. To correctly conclude the exercise, we should think about what the most important biological features are (like if all the steps in the chain must be considered or if other inputs might play a role, etc.). We mentioned the delay, but perhaps there are other features that the model must be able to simulate. This is when biological insight comes into play.

The real model used to generate the data was a third-order model, and included nonlinearities and delays (see appendix). For these reasons, it was not possible to find the exact parameter values of the original system using a linear model. However, the obtained models (2nd or 3rd-order models) were reasonably close in terms of model order and could simulate the main features of the data.

### 9.2.6 Limitations of System Identification

System identification, when used carefully, can be an extremely powerful technique to elucidate the dynamics of a system. We would like to conclude this tutorial by giving some advice and warnings on the limitations of this method. We stress the fact that a model should be built from the information contained in the data and in any prior knowledge of the system. It is essential to have very informative data that covers as many experimental conditions as possible. Next we explain what very informative data mean.

In the frequency domain, a signal is decomposed into a sum of sine waves each with a different period. If the frequency is more relevant than others, then stimulating the system with the corresponding sine wave will produce a larger response. Naturally, a model will be more reliable at the frequencies that comprise the main components of the input signal used in the estimation process. Consequently, the reader should be

careful in choosing the dataset for the estimation process. For example, the response of a system to a periodic input is likely to provide information only around its specific period; a model obtained from this data is likely to have a very low predictive power away from that frequency.

Model selection favours simplicity. If a low order model has a good fit with the data and can reproduce the qualitative features of the validation set, then the model estimation process can stop. Increasing the model order should only be done to achieve a better qualitative simulation. Trying to achieve a perfect fit will likely lead to over-fitting instead of improvement, and it risks obtaining a model that reproduces the data but not the real dynamical system.

In previous sections, we introduced the problem of local minima. Noisy signals or high order models can result in cost functions with several local minima, which cause optimisation algorithms to struggle to converge to an interesting solution. Also forcing a structure on the model might have this result.

Finally, most systems in real life are nonlinear. Thus a natural objection to linear modelling is that it inherently can't describe aspects of the system. However, depending on the particular application, linear systems have demonstrated to be very good approximations of nonlinear systems. In those cases, and given their relatively simplicity, it makes sense to prefer this class of systems to nonlinear systems if they can provide good predictions.

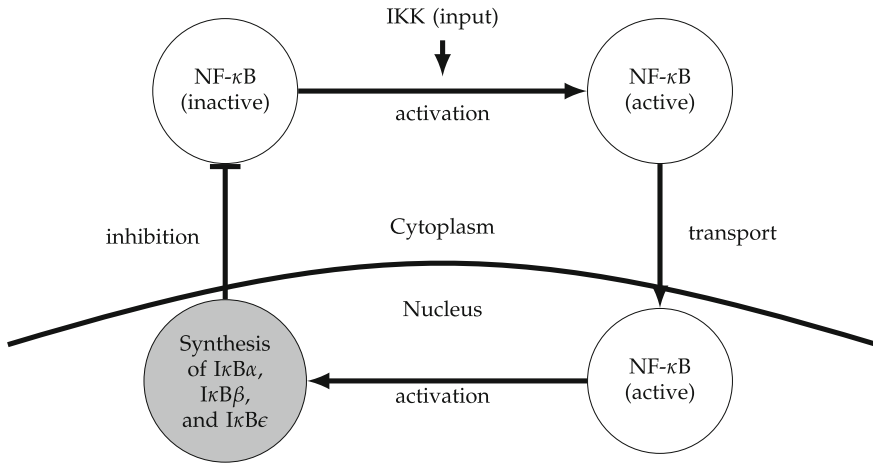
## 9.3 Biological Examples of Successful Application

To conclude this chapter, we will present two examples of successful applications in the literature of linear system identification. The first is based on the paper 'Achieving Stability of Lipopolysaccharide-Induced NF- $\kappa$ B Activation' [3] from the Baltimore laboratory (California Institute of Technology), and the second is based on the paper 'Correct Biological Timing in Arabidopsis Requires Multiple Light-Signalling Pathways' [4] from the laboratory (University of Cambridge, UK).

### 9.3.1 Modelling the NF- $\kappa$ B Signalling Pathway

The NF- $\kappa$ B signalling pathway is a key process for gene regulation in inter- and intracellular signalling, cellular stress responses, cell growth, survival, and apoptosis. Deciphering its temporal and specificity control is therefore of the utmost importance for a better understanding of cell physiology. The Baltimore lab has been working on this pathway for several years and the paper [3] is an example of how mathematical modelling can be used to speed up the process of experimental science.

In mammals, NF- $\kappa$ B is a transcription factor that is involved in multiple regulatory pathways. It is involved in metabolic processes like inflammatory responses, immune system development, apoptosis, learning in the brain, and bone development. Aber-



**Fig. 9.15** Feedback of the NF-κB signalling pathway [6]

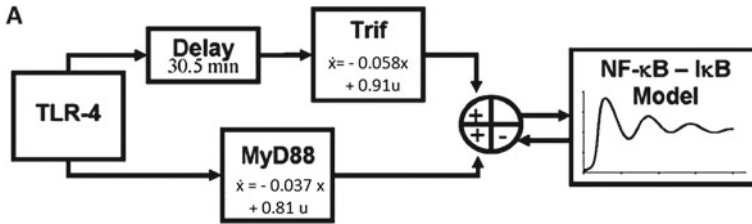
rant NF-κB activity has been linked to oncogenesis, tumour progression, and resistance to chemotherapy. Understanding NF-κB activation is therefore very important in cancer research. A previous study by Baltimore’s lab [6] gives a description of how the IκB-NF-κB signalling module might work. Using a combination of computational models and experiments, they elucidated the strong negative feedback loop involving the protein complex IκB that allows a fast turn-off of the NF-κB response. IκB holds NF-κB inactive in the cytoplasm until IκB is degraded by the IKK complex, which is activated by cell stimulation through TNFα expression. NF-κB is then free to translocate into the nucleus where it activates several pathways, including synthesis of IκB proteins (IκBα, IκBβ and IκBε) that in turn control NF-κB activation. This feedback is represented in Fig. 9.15.

NF-κB shows damped oscillations as a result of this regulation. In [6], the feedback pathway was represented with the following model:

$$\frac{dx}{dt} = S - \alpha x - \beta y \tag{9.22}$$

$$\frac{dy}{dt} = \gamma x - \delta y \tag{9.23}$$

where  $S$  represents the stimulus (the input). The model structure tells us immediately that we are describing a feedback system, since both differential equations depend on both  $x$  and  $y$ . The strength of the feedback is determined by the parameters  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ . In particular, two parameters determine the damping of the system: large oscillations (high feedback power) corresponds to low values of  $\alpha$  and  $\delta$ , and a quasi steady state behaviour (high damping) corresponds to high values of  $\alpha$  and  $\delta$ . These parameters were estimated using a combination of known association rates and the application of system identification to experimental data.



**Fig. 9.16** Feedback system of the TLR-4 pathways to NF- $\kappa$ B expression [3]. Figure taken from [3], with permission

Following this first model, new components of this feedback regulation were discovered and a more complete description was sought. The starting point of this new study was that cells stimulated with lipopolysaccharide (LPS) showed non-oscillatory dynamics with respect to active NF- $\kappa$ B. Therefore LPS must be involved in its regulation.

LPS activates expression of the gene TLR4, which has two downstream pathways, both of which regulate NF- $\kappa$ B. One pathway is MYD88 dependent, and it has been almost completely described as an activator of IKK synthesis. The other pathway had not yet been fully understood, but it has the same end result of degrading I $\kappa$ B in the TNF $\alpha$ -activated pathway and acting through an adaptor called Trif.

Analysis of experimental results showed that the MYD88-dependent pathway occurs earlier than the independent one. If both pathways are inactive (MYD88 and Trif-null double mutant), there is no NF- $\kappa$ B activation, while each single null-mutation results in NF- $\kappa$ B oscillations. The observed NF- $\kappa$ B oscillations could be a result of an interaction between these two pathways.

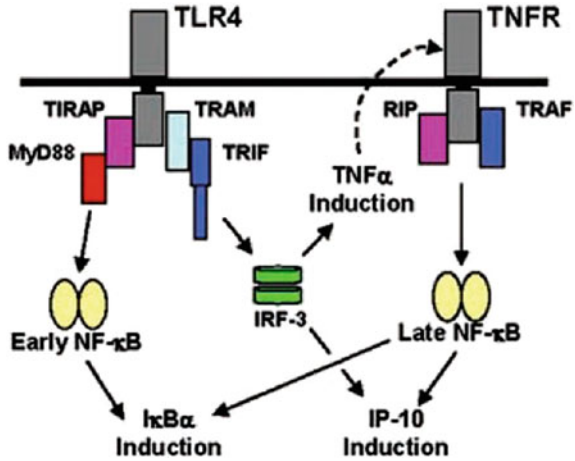
Covert et al. [3] tested the hypothesis that NF- $\kappa$ B oscillations are a result of the IKK complex being activated at different times in each pathway. Either the two pathways share similar kinetics, which prevents them from progressing simultaneously, or the MYD88-independent pathway requires more steps.

To rule out one of the two possibilities, linear models representing each of the two pathways were built and added to the previous model of the NF- $\kappa$ B-I $\kappa$ B signalling module in a feedback loop. The two models were estimated using a first order structure and empirically determined protein concentration as driving dataset.

The model shown in Fig. 9.16 points out that the kinetics of the two linear models are indeed very similar (by simple comparison of the coefficients) but that a delay of about half a hour is required for the Trif pathway to reproduce the data correctly.

Being the minimal model (1st order), some predictions are not confirmed in the experimental process (for instance, discrepancies in the period for I $\kappa$ B $\alpha$  protein synthesis). However, the predictive power of the model is quite significant as it correctly predicts oscillations of I $\kappa$ B $\alpha$  protein level in MYD88-null and Trif-null mutants, but not in the wild-type. Hence, they tested the hypothesis of a longer kinetic pathway as cause of the delay in the Trif-dependent signalling. Experimental

**Fig. 9.17** Proposed pathway of activation of NF- $\kappa$ B through TLR4 control. Figure taken from [3], with permission



evidence shows that the MYD88-independent pathway does indeed require protein synthesis (longer kinetic).

They isolated the components of the Trif pathway, finding the transcription factor IRF3 and the known protein TNF $\alpha$  to be part of the down-regulated cascade (see Fig. 9.17). It is compelling that a very basic model (1st order linear system with a delay) can provide information on a complex multi-input system. The underlying biological system undoubtedly incorporates nonlinear behaviours and hence cannot be completely represented by a linear model. However, a simple LTI scheme captures its most important features, the delay in the MYD88-independent pathway. This shows that additional complexity, in general harder to characterise, is also not required to explain the interdependence of I $\kappa$ B and NF- $\kappa$ B signals.

### 9.3.2 Regulation of $[Ca^{2+}]_{cyt}$ by Light and the Circadian Clock in *Arabidopsis*

Circadian clocks confer the ability of an organism to align its physiology with the daily rotation of the planet, which leads to 24 h periodic cycles in light availability and temperature. This ability is a result of genetic networks that generate autonomous oscillations and provide rhythmic cues to downstream gene expression and signalling. *Arabidopsis thaliana* is a model organism for plant biology because of its relatively small genome and short lifecycle. The circadian clock of *A. thaliana* has been studied closely and many of its features are well understood. A number of essential clock components have been identified, including CIRCADIAN CLOCK ASSOCIATED 1 (CCA1), LATE ELONGATED HYPOCOTYL (LHY), TIMING OF CAB2 EXPRESSION 1 (TOC1) and GIGANTEA (GI), which all play a role in sustaining circadian oscillations [8]. Both experimental and computational analyses have driven

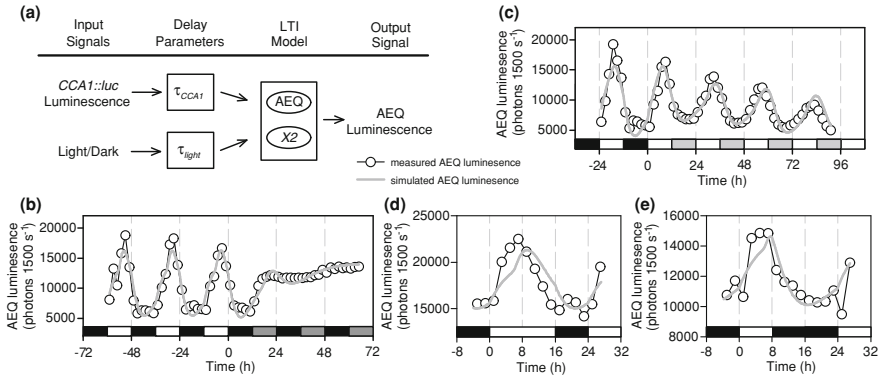
progress in uncovering the interactions between clock-associated genes that leads to robust timekeeping in *A. thaliana*. Recent computational model of the clock likens the genetic network to the classical repressilator mechanism, incorporating multiple negative feedback loops of transcriptional control [11].

In this section, we will describe our previous work that used system identification to understand how circadian clocks and light signalling pathways contribute to the regulation of physiology in *A. thaliana* [4]. The *external coincidence* hypothesis proposes that the phase of a circadian regulated gene is the result of a coincidence between the phase of the main oscillator and the light/dark cycle. On the other hand, the *internal coincidence* model proposes that light entrains two different rhythms in the main oscillator (morning and evening loops) and the relationship between the two defines the phase of the output.

In order to test the two hypotheses, we sought to understand and quantify the regulation of the concentration of cytosolic-free  $\text{Ca}^{2+}$  ( $[\text{Ca}^{2+}]_{\text{cyt}}$ ), an important signalling ion in cellular organisms. Previous experimental analyses had shown that  $[\text{Ca}^{2+}]_{\text{cyt}}$  is regulated by both the circadian oscillator and light signalling and that the phase of circadian  $[\text{Ca}^{2+}]_{\text{cyt}}$  oscillations changes in response to photoperiod [12]. Moreover, plants lacking CCA1 had no circadian oscillations of  $[\text{Ca}^{2+}]_{\text{cyt}}$ , despite there being (albeit short-period) oscillations in other clock outputs. As no other essential biochemical components involved in regulating  $[\text{Ca}^{2+}]_{\text{cyt}}$  had been identified, we constructed a model of  $[\text{Ca}^{2+}]_{\text{cyt}}$  with two inputs: light and CCA1 expression (Fig. 9.18a). This corresponds to the incorporation of light at two levels, as CCA1 expression is itself regulated by light. Since the signalling pathways linking the inputs to  $[\text{Ca}^{2+}]_{\text{cyt}}$  were not well understood, several model orders were compared. Moreover, varying timescales in the regulation by CCA1 and light were considered by introducing delay parameters, a method that helps reduce model order as simple delays can account for the internal complexity of each state variable.

Estimation was done using data from a single experiment, where input (CCA1) and output ( $[\text{Ca}^{2+}]_{\text{cyt}}$ ) measurements were collected in 12 h light/12 h dark cycles (12L/12D) followed by an extended period of constant dark (Fig. 9.18b). This was a dynamically useful choice of input as it corresponds to a square wave followed by a step function, and consequently excites all frequencies. Validation was then conducted on three datasets with different light/dark conditions to assess model performance (Fig. 9.18c–e). A weighted correlation metric was used to determine the optimal model order in terms of the predictive ability of the model, which suggested a necessary role for a single hidden variable (hereafter referred to as  $X2$ ). The selected model successfully predicted the behaviour of the CCA1 null mutation against experimental data, providing further support for the model and the LTI system identification approach (Fig. 9.19a).

As the purpose of system identification was to elucidate the signalling pathway leading up to  $[\text{Ca}^{2+}]_{\text{cyt}}$  the next step was to understand the role of  $X2$  in the model. A mutation in the  $X2$  component was simulated in 16L/8D cycles and compared with experimental data measuring a variety of genetic mutants in corresponding conditions. A close match was observed with the PHYTOCHROME A (PHYA) mutant, *phyA-201* (Fig. 9.19c). PHYA is a red light photoreceptor and mediates far

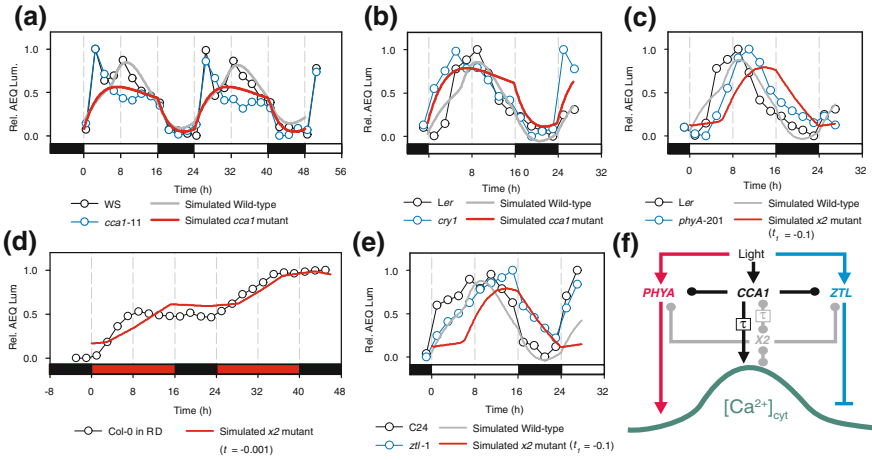


**Fig. 9.18** Constructing an LTI model for the circadian regulation of  $[Ca^{2+}]_{cyt}$ . **a** Schematic for the model structure; **b** simulation of the estimation dataset using the selected model; **c** model validation using  $[Ca^{2+}]_{cyt}$  expression measured in 12L/12D and then constant light; **d**, **e** model validation using  $[Ca^{2+}]_{cyt}$  expression measured in light condition 16L/8D and 8L/16D, respectively. This figure is reproduced from [4]

red light signals to the clock [10]. However, PHYA was not the only candidate for X2. Blue light seems to be required for the decrease of  $[Ca^{2+}]_{cyt}$  at the end of the light period, as in cycles of red-light and dark,  $[Ca^{2+}]_{cyt}$  progressively increases during each circadian period (Fig. 9.19d), as opposed to the stable oscillations observed in white light cycles. Simulations of the *x2* mutant showed similarities to experimental data of the blue light receptor mutant *ztl-1* (Fig. 9.19e, suggesting that X2 may represent more than one biochemical component (Fig. 9.19f). We concluded that X2 might represent the circadian regulation of light signalling pathways as it describes time dependent effects of light input.

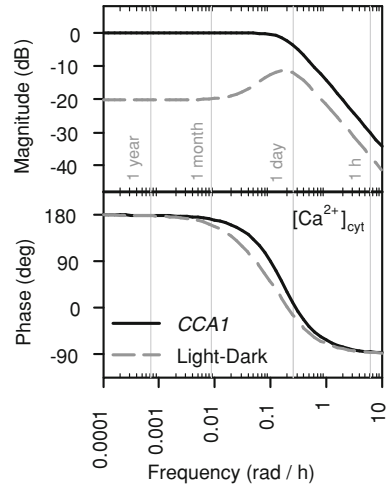
To investigate the importance of timing of the inputs, Bode analysis was applied to the model (Fig. 9.20). In order to use a Bode plot, the input is decomposed into a sum of sinusoidal signals with different periods. Each of these periodic signals causes a different response in the output. Each period is then plotted against its unique response. In this case, it was found that the CCA1 input dominates over the light signal at almost all frequencies but especially at lower ones (low frequencies corresponds to long period oscillations). This suggests that light input only modulates  $[Ca^{2+}]_{cyt}$  over faster variations in light availability.

To determine whether rapid light input regulation could be a more general phenomenon, the whole rhythmic transcriptome (3,503 genes) of *A. thaliana* was explored using LTI models and Bode analysis. Models were estimated using CCA1 or TOC1 and light availability as inputs, and published microarrays as driving data. Validation was then conducted by evaluating a combination of fitness over multiple datasets (1,083 models using CCA1 and 460 using TOC1). CCA1 proved to be a more suitable driver for the models, with the majority of TOC1-driven models also performing well with CCA1 as input. We conducted Bode analysis on the resulting models. Magnitude plots that resembled (same or smaller difference in magnitude between the two



**Fig. 9.19** Comparison of simulated mutations with experimental data. Perturbations to the second order LTI model of  $[Ca^{2+}]_{cyt}$  simulate mutations in the pathway components. **a** Comparing simulated *cca1* mutation with *cca1-11* null mutant data. **b** Comparing simulated *cca1* mutation with *cry1-1* mutant data. **c** Comparing simulated *x2* mutant with *phyA-201* mutant data. **d** Comparing simulated *x2* mutant with wild-type data in cycles of red light and dark. **e** Comparing simulated *x2* mutant with *ztl-1* mutant data. **f** Schematic proposing the relationship between biochemical components in the regulation of  $[Ca^{2+}]_{cyt}$

**Fig. 9.20** Bode plot of the two inputs model. At each frequency, magnitude (*top figure*) represents the amplitude of the output signal, while phase (*bottom figure*) expresses the difference in oscillation phase of the input and output signals



inputs pathways) Fig. 9.20 were classified as being coregulated by light and CCA1. Light- or clock-dominated regulation was defined for higher magnitude differences. Our results agree with experimental evidence, showing that genes like LHY, PRR5/7 or GI are coregulated by light and the clock. We compared model classes and the phase of peak transcript abundance in daily regimes with long photoperiods (16L/8D)



and short photoperiods (8L/16D). We concluded that clock-dominated transcripts are associated with morning expression independently of the photoperiod, while light-dominated and coregulated models have altered peak times and generally peak later in the day/night.

Using system identification applied to LTI models, and tools from Control theory, it was possible to understand the timing of the regulation, its frequency response, and predict biochemical components and their specific roles in mediating daily control of  $[Ca^{2+}]_{\text{cyt}}$ . We demonstrated that the observed oscillations of  $[Ca^{2+}]_{\text{cyt}}$  are a result of a combination of rapid light signals and circadian inputs. As the two inputs act at different timescales, and are mediated by a single autonomous oscillator, this case study supports the external coincidence hypothesis. Moreover, the significant speed of this system identification technique allowed us to perform genome-scaled analysis and to give a complete characterisation of the whole rhythmic genome.

## 9.4 Summary

This chapter explained the main ideas behind SI and showed how to construct a model from input/output data. The process of model estimation is the first step of the machinery. We showed how this can be achieved by minimising a cost function using an optimisation algorithm. Prior knowledge of the system can be incorporated in the process by constraining specific structures and/or initial conditions. The optimisation algorithm is the weak point of this procedure as getting stuck in local minima could be a major limitation. In order to be reliable, estimated models need to be flexible and correctly reproduce different datasets. This is addressed in the model validation procedure. The model is tested with a new input/output set from the same system. This new performance should be comparable in terms of fitness to the one obtained with the estimation dataset. Model validation is fundamental to prevent over-fitting, which can sensibly reduce the model predictive power.

Once models of different orders have been estimated and validated, we need to select one as our best representation of the system. There are several criteria that can be used for model selection. We have seen fitness comparison and the Akaike Information Criteria. We also mentioned how some qualitative biological features can be used as thresholds. When the models performs approximately equally, then the simplest model has to be preferred.

Finally, we discussed some of the limitation of SI. Local minima, over-fitting and nonlinearities are the main drawbacks of using LTI models. The key idea is that a good model extracts as much information as possible from data. Informative non-noisy data are more likely to result in reliable models, while sparse and corrupted signals produce models with limited predictive power and unfit for further analysis.

## 9.5 Model Used in Simulations

The dataset of progesterone and cholesterol concentrations introduced in this chapter were simulated using the following model:

$$\frac{dx_1}{dt} = -5.3x_1(t - 0.05) + 3.1x_2(t - 0.05) + 2.7u(t - 0.05) \quad (9.24)$$

$$\frac{dx_2}{dt} = 5.3x_1(t - 0.05) - 7.9x_2(t - 0.05) + 2.7x_3(t - 0.05) + \frac{1}{5}e^{-x_2(t-0.05)} \quad (9.25)$$

$$\frac{dx_3}{dt} = 4.8x_2(t - 0.05) - 2.7x_3(t - 0.05) \quad (9.26)$$

This model has 3 states, 2 representing progesterone and cholesterol concentration and one representing the internal steps of the process. This additional state is a hidden variable and explains why a 2nd order model description wasn't enough to describe the data.

This model is not meant to represent every single detail of the steroidogenesis process, but instead to capture some of its interesting features and with enough dynamic behaviour to illustrate the system identification features. Here, delays represent either unknown pathways and/or delays in the reaction time, while nonlinearities describe fast activation processes. The initial conditions used to generate the data for the noise-free and the noisy estimation in Figs. 9.2 and 9.11 are (10, 1, 1). The noise was drawn from a random normal distribution with mean 0 and standard deviation 1/10.

## References

1. Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19(6):716–723
2. Bishop C et al (2006) *Pattern recognition and machine learning*, vol 4. Springer, New York
3. Covert MW, Leung TH, Gaston JE, Baltimore D (2005) Achieving stability of lipopolysaccharide-induced NF- $\kappa$ B activation. *Science* 309(5742):1854–1857. doi:10.1126/science.1112304, <http://dx.doi.org/10.1126/science.1112304>
4. Dalchau N, Hubbard K, Robertson F, Hotta C, Briggs H, Stan G, Gonçalves J, Webb A (2010) Correct biological timing in arabidopsis requires multiple light-signaling pathways. *Proc Natl Acad Sci USA* 107(29):13171–13176. doi:10.1073/pnas.1001429107, <http://dx.doi.org/10.1073/pnas.1001429107>
5. Gibson S, Ninness B (2005) Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica* 41(10):1667–1682
6. Hoffmann A, Levchenko A, Scott M, Baltimore D (2002) The  $\kappa$ b-nf- $\kappa$ b signaling module: temporal control and selective gene activation. *Science* 298(5596):1241–1245
7. Ljung L (1999) *Systems identification: theory for the user*. Prentice Hall, Englewood
8. Nagel DH, Kay SA (2012) Complexity in the wiring and regulation of plant circadian networks. *Curr Biol* 22(16):R648–R657. doi:10.1016/j.cub.2012.07.025, <http://dx.doi.org/10.1016/j.cub.2012.07.025>

9. Nelson D, Cox M (2008) *Lehninger principles of biochemistry*. Wh Freeman, New York
10. Parks B, Quail P, Hangarter R (1996) Phytochrome a regulates red-light induction of phototropic enhancement in *Arabidopsis*. *Plant Physiol* 110(1):155–162
11. Pokhilko A, Fernández AP, Edwards KD, Southern MM, Halliday KJ, Millar AJ (2012) The clock gene circuit in *Arabidopsis* includes a repressilator with additional feedback loops. *Mol Syst Biol* 8:574. doi:[10.1038/msb.2012.6](https://doi.org/10.1038/msb.2012.6), <http://dx.doi.org/10.1038/msb.2012.6>
12. Xu X, Hotta CT, Dodd AN, Love J, Sharrock R, Lee YW, Xie Q, Johnson CH, Webb A (2007) Distinct light and clock modulation of cytosolic free Ca<sup>2+</sup> oscillations and rhythmic chlorophyll a/b binding protein2 promoter activity in *Arabidopsis*. *Plant Cell* 19(11):3474–3490. doi:[10.1105/tpc.106.046011](https://doi.org/10.1105/tpc.106.046011), <http://dx.doi.org/10.1105/tpc.106.046011>

# Chapter 10

## Model Checking in Biology

Jasmin Fisher and Nir Piterman

**Abstract** Model checking is a technique to check whether programs and designs satisfy properties expressed in temporal logic. Such properties characterize sequences of events. In recent years, model checking has become a familiar tool in software and hardware industries. One of the main strengths of model checking is its ability to supply counter examples: in case that the property is not satisfied by the model we get an execution exhibiting this failure. Counter examples are fundamental in understanding, localizing, and eventually fixing, faults. This, together with the relative ease of use of model checking, led to its adoption. The success of model checking prompted system biologists to harness it to their needs. In this domain, the main usage is to have a model representing a certain biological phenomenon and to use model checking for one of two things. Either prove that the model satisfies a set of properties, i.e., reproduces a set of biological behaviors. Or to use model checking to extract interesting behaviors of the model by looking for a counter example to the property saying that this interesting behavior does not happen. In this chapter we present the technique of model checking and survey its usage in systems biology. We take quite a liberal interpretation of what is model checking and consider also cases where the techniques underlying model checking are used for similar purposes in systems biology.

**Keywords** Model checking · Temporal logic · Transition system · Path · Computation · Transition · Computation tree logic (CTL) · Linear temporal logic (LTL) · Unary operator · Binary operator · Boolean operator

---

J. Fisher (✉)

Microsoft Research Cambridge, Cambridge, UK  
e-mail: jasmin.fisher@microsoft.com

N. Piterman

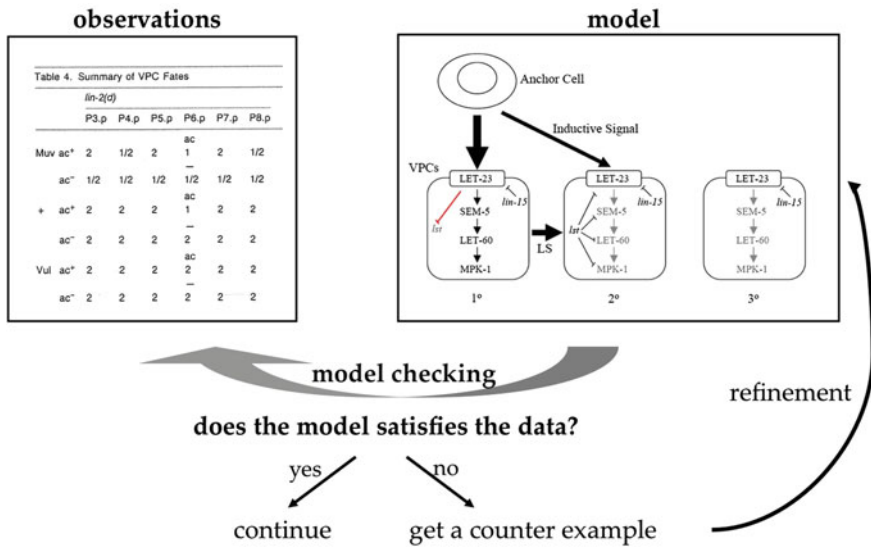
University of Leicester, Leicester, UK  
e-mail: nir.piterman@leicester.ac.uk

## 10.1 Introduction

Biological systems are extremely complex reactive systems. They operate as highly concurrent programs with millions of entities running in parallel and communicating with each other under various environmental conditions. Understanding how living systems operate in such harmony and precision, and how this harmony is being disrupted in disease states, are key questions in biological and medical research. Due to their enormous complexity, the comprehension and analysis of living systems is a major challenge. Over the last decade various efforts to tackle this problem of enormous complexity concentrate on a new approach called *Executable Biology* focused on the construction and analysis of executable models describing biological phenomena (for a review see [19]).

Over the years, these efforts have demonstrated successfully how the use of formal methods can be beneficial for gaining new biological insights and even directing new experimental avenues. At the core of these models lies their ability to be analysed by *model checking* [15]. In the context of biological models, model checking can be used in two ways:

1. **Testing and comparing hypotheses.** Computational models represent hypotheses about molecular and cellular mechanisms that result in experimental data. Executions of these models can be used to check if a possible outcome of these mechanisms conforms to the data. Due to the nondeterministic nature of these models, each repeated execution may yield a different possible outcome. Therefore it is impossible to check by executing these models if all possible outcomes conform to the data. This, however, can be done by model checking, as model checking systematically analyzes all of the infinitely many possible outcomes of a computational model without executing them one by one. If model checking tells us (a) that all possible outcomes of the computational model agree with the experimental data, and (b) that all experimental outcomes can be reproduced by the model, then the model represents a mechanism that explains the experimental data. If (b) is violated, then the hypothesis that the computational model captures a mechanism for explaining the data is found to be wrong. In this case, either the model must be enriched as to produce the additional outcomes that are present in the data, or completely revised. If (a) is violated, then the situation is more interesting. In this case, the mechanistic hypothesis represented by the model may be wrong, and one may attempt to restrict the model as to not produce outcomes that are not supported by the data. Alternatively, the experimental data may be incomplete and not exhibit some possible observations that would show up if more data were collected. Thus, in case (a), model checking can offer suggestions for additional, targeted experiments that would either confirm or invalidate the mechanistic hypothesis represented by the computational model (Fig. 10.1).
2. **Querying the behaviour of mechanistic models.** Once a model has been tested and compared against hypotheses, it can also be queried by searching for interesting executions. By stating that a certain property holds for all executions, or by stating that a certain property does not hold for all executions, we can either



**Fig. 10.1** Methodology of using model checking. One possible methodology for using model checking is by comparing mechanistic models to specifications. A formal model that represents a hypothetical understanding of the system under study is constructed (*model*). Results of experiments are formalized in the form of specifications (*observations*). Model checking is used to ensure that the model reproduces the experimental observations. Mismatch with experimental observations suggests that the model is lacking and should be refined by additional information. Match with experimental observations could lead to further querying and testing of the model to suggest further experimental studies

validate or falsify specific predictions about the behaviour of the model. By phrasing queries such as which molecular events may lead to specific cell behaviour, we can also determine what part of the execution allows this kind of events.

In this chapter we give an introduction to model checking and the techniques underlying it. This is in the hope that practitioners will understand better the techniques and what can be done with them. In particular, recent algorithmic progress shows that with good understanding of the underlying techniques further types of analysis can be accomplished using model checking techniques. We also give examples of instances where model checking was used in biological modeling to demonstrate a flavor of the results that can be achieved by using model checking. In particular, in some cases, usage of model checking led to new biological insights, shedding new light on signalling crosstalk.

## 10.2 What Is Model Checking?

In this section we introduce the mathematical concepts underlying model checking. Model checking is a technique that checks whether all the computations of a system satisfy a property. In order to be able to answer this question we have to create a formal

model of the system and formally state the property. In this section we introduce the underlying formal concepts and the formal definition of model checking. We start by introducing transition systems, which will be used to represent the possible computations of a model. We then proceed to explain temporal logic and conclude with an explanation of what is model checking.

### 10.2.1 Transition Systems

The concept of a transition system is a fundamental concept of computation. Here, we are going to refer to a transition system representing some machine. However, in the context of this chapter, the machine is usually a model of some biological machinery. A transition system has states, which represent snapshots of the status of the machine, and transitions, which represent the possible change of status of the machine. For the sake of model checking additional annotations are required and these are usually put on the states in the form of propositions, which are basic facts about the world of the machine that could be either true or false in a given state.

A transition system is  $\mathcal{T} = (S, T, S_0, \mathcal{P}, L)$  with the following components:

- $\mathcal{T}$  is the name of the transition system.
- $S$  is the set of states of  $\mathcal{T}$ , every state  $s \in S$  represents a possible snapshot of the machine. The set  $S_0$  is a subset of  $S$  of initial states indicating in which states can the machine start an execution. For the purpose of presentation of model checking we are going to concentrate on transition systems with a finite number of states.
- $T$  is the set of transitions formally represented as a set of pairs of states, i.e.,  $T \subseteq S \times S$ . For states  $s$  and  $t$ , having  $(s, t) \in T$  means that the machine can change from state  $s$  to state  $t$ .
- $\mathcal{P}$  is the set of facts that are observable in a state of  $\mathcal{T}$ . The labeling function  $L$  associates with every state the facts that are true in that state. Formally,  $L: S \rightarrow 2^{\mathcal{P}}$  is a function that associates with every state the set of propositions that are true in it. By elimination, all other propositions are false in that state.

Given this notion of a transition system we are now ready to define what are computations of a transition system. Intuitively, a computation is a sequence of states that starts in an initial state and proceeds by taking permissible transitions. More formally, a path  $\pi$  is a sequence  $s_0, s_1, \dots$  such that all transitions are in  $T$ , i.e., for every  $i \geq 0$  we have  $(s_i, s_{i+1}) \in T$ . A path  $\pi = s_0, s_1, \dots$  is a computation if in addition  $s_0$  is initial, i.e.,  $s_0 \in S_0$ . We note that paths and computations are infinite. Thus, we assume that all states have at least one outgoing transition. This is a usual assumption in model checking as it makes it simpler to avoid many technicalities. One can reintroduce the concept of finite runs by adding a special *finish* state with a self loop and considering runs that end in an infinite sequence of visits to finish as finite. A computation or a path induces a sequence of labels representing the change in the status of the different facts about the machine that interest us. Instead of looking on the sequence  $s_0, s_1, \dots$ , we could look on the sequence of labels  $L(s_0), L(s_1), \dots$ , where  $L(s_i)$  is the set of facts that are true in state  $s_i$  (and the rest are false). We call

the sequence of labels a *run* and denote it by  $r = L(s_0, s_1, \dots)$ . When the run is induced by a computation we say that it is *initialized*.

### 10.2.2 Temporal Logic

We now turn to the definition of a languages to give “interesting” properties about the transition system. Essentially, we would like to be able to describe qualities of the transition system or its computations. The ultimate goal (to be visited in the next subsection) is to check whether a transition system satisfies a given property. That is, whether the machine described by a transition system has this quality or not. We distinguish between two languages to describe properties of systems. The first, called *linear temporal logic* views a system as the sum of all of its possible computations. The second, called *computation tree logic* views the entire transition system as the embedding of the machine. Here we choose to expose both as each has its own advantages and both have been used in the context of biological modeling.

We start with *linear temporal logic* (LTL for short), which was introduced by Pnueli in the late 1970’s [29]. As it’s name suggests it takes the first approach of viewing the system as the sum of all its computations. An LTL formula is going to use the basic facts about states (i.e., labels) and combine them in ways that say things about sequences. Then, we define when an LTL formula is satisfied by a single run. Finally, an LTL formula will be satisfied by the transition system if all possible runs of the transition system satisfy it.

As mentioned an LTL formula can use one of the basic facts in P. It can use one of the following operators.

- Next, denoted  $\mathcal{X}$ , a unary operator saying that its operand is true in the next state.
- Until, denoted  $\mathcal{U}$ , a binary operator saying that its first operand must hold until its second operand holds.
- Always (or globally), denoted  $\mathcal{G}$ , a unary operator saying that its operand is true in all future states (including current).
- Eventually (or finally), denoted  $\mathcal{F}$ , a unary operator saying that its operand is true in some future (or current) state.
- In addition, LTL uses the usual Boolean operators not, denoted  $\neg$ , conjunction, denoted  $\wedge$ , disjunction, denoted  $\vee$ , and implication, denoted  $\rightarrow$ .

Thus, an LTL formula is constructed hierarchically from simpler formulas. For example, the formula  $\mathcal{G}(p \rightarrow \mathcal{X}q)$  uses the basic facts  $p$  and  $q$  and says that in all states of the computation if  $p$  holds in a state then  $q$  must hold in the next state. Similarly,  $\mathcal{G}(p \rightarrow p\mathcal{U}q)$  says that whenever  $p$  holds, it must hold continuously until a later time when  $q$  holds.

We now make the intuition regarding when a formula holds in a computation formal. For that, we start with a definition of when a formula holds in a specific run in a specific location. The definition builds truth values according to the hierarchical structure of the formula. That is, basic facts (propositions) can be deduced from



the label. Then, the truth of more complicated formulas is constructed from that of simpler formulas. Consider a run  $r = l_1, l_2, \dots$  over  $\mathcal{P}$ . That is, every  $l_i$  is a set of basic facts that are true at time  $i$  (and all facts in  $\mathcal{P} - l_i$  are false at time  $i$ ). The following list defines when an LTL formula  $\varphi$  is satisfied in  $r$  at time  $i$ , denoted  $r, i \models \varphi$ , and when it is not satisfied, denoted  $r, i \not\models \varphi$ .

- If  $\varphi$  is a proposition, then  $r, i \models \varphi$  if  $\varphi \in l_i$  and  $r, i \not\models \varphi$  if  $\varphi \notin l_i$ .
- If  $\varphi$  is  $\neg\psi$  then  $r, i \models \varphi$  if  $r, i \not\models \psi$  and  $r, i \not\models \varphi$  if  $r, i \models \psi$ .
- If  $\varphi$  is  $\psi_1 \wedge \psi_2$  then  $r, i \models \varphi$  if  $r, i \models \psi_1$  and  $r, i \models \psi_2$  and  $r, i \not\models \varphi$  if either  $r, i \not\models \psi_1$  or  $r, i \not\models \psi_2$ .
- If  $\varphi$  is  $\psi_1 \vee \psi_2$  then  $r, i \models \varphi$  if either  $r, i \models \psi_1$  or  $r, i \models \psi_2$  and  $r, i \not\models \varphi$  if  $r, i \not\models \psi_1$  and  $r, i \not\models \psi_2$ .
- If  $\varphi$  is  $\psi_1 \rightarrow \psi_2$  then  $r, i \models \varphi$  if either  $r, i \not\models \psi_1$  or  $r, i \models \psi_2$  and  $r, i \not\models \varphi$  if  $r, i \models \psi_1$  and  $r, i \not\models \psi_2$ .
- If  $\varphi$  is  $\mathcal{X}\psi$  then  $r, i \models \varphi$  if  $r, i + 1 \models \psi$  and  $r, i \not\models \varphi$  if  $r, i + 1 \not\models \psi$ .
- If  $\varphi$  is  $\psi_1 \mathcal{U} \psi_2$  then  $r, i \models \varphi$  if there is a  $k \geq i$  such that  $r, k \models \psi_2$  and for every  $i \leq j < k$  we have  $r, j \models \psi_1$  and  $r, i \not\models \varphi$  if for every  $k \geq i$  such that  $r, k \models \psi_2$  there is  $i \leq j < k$  such that  $r, j \not\models \psi_1$ .
- If  $\varphi$  is  $\mathcal{G}\psi$  then  $r, i \models \varphi$  if for every  $j \geq i$  we have  $r, j \models \psi$  and  $r, i \not\models \varphi$  if for some  $j \geq i$  we have  $r, j \not\models \psi$ .
- If  $\varphi$  is  $\mathcal{F}\psi$  then  $r, i \models \varphi$  if for some  $j \geq i$  we have  $r, j \models \psi$  and  $r, i \not\models \varphi$  if for all  $j \geq i$  we have  $r, j \not\models \psi$ .

Finally, a system  $\mathcal{T}$  satisfies an LTL formula  $\varphi$ , denoted  $\mathcal{T} \models \varphi$ , if for every run of the system we have  $r, 0 \models \varphi$ . Otherwise, the system does not satisfy the formula, denoted  $\mathcal{T} \not\models \varphi$ .

We turn now to *computation tree logic* (CTL for short), which was introduced by Clarke and Emerson [14]. The name of the logic derives from viewing the transition system as producing a single *computation tree*, which we do not explain here. Unlike LTL, CTL is going to take an integrative view of the system. A formula is going to be either true or false for a state of the system. Like LTL, CTL is going to use the basic facts about states and combine them to properties about the system. It then combines information about the system by considering the paths that start in states and state properties of some or all these paths. A CTL formula is satisfied by the system if all initial states of the system satisfy it.

As mentioned CTL combines information about paths and about states. A CTL formula can use one of the basic facts in  $\mathcal{P}$ . Such basic facts are state formulas. It can use one of the following operators.

- Boolean operators  $\wedge, \vee, \neg, \rightarrow$  can be applied to formulas as in LTL.
- The temporal operators next, until, always, and eventually are combined with path quantification  $E$  and  $A$ . Thus, CTL includes the unary operators  $E\mathcal{X}, A\mathcal{X}, E\mathcal{G}, A\mathcal{G}, E\mathcal{F},$  and  $A\mathcal{F}$  that can be applied to simpler formulas. The binary operators  $E\mathcal{U}$  and  $A\mathcal{U}$  combine two formulas. The  $E$  quantifier says “there exists a path” and the  $A$  quantifier says “for all paths”. The meaning of the temporal part remains the same as in LTL. Thus, a formula like  $A\mathcal{X}\psi$  says that all next states satisfy

the property  $\psi$ . A formula like  $E(\psi_1 \mathcal{U} \psi_2)$  says that from a given state there is a path satisfying the property  $\psi_1 \mathcal{U} \psi_2$ .

For example, the formula  $A\mathcal{G}(p \rightarrow E\mathcal{X}q)$  says that every state where the fact  $p$  is true that is reachable from an initial state must have a successor where the fact  $q$  holds. Similarly,  $A\mathcal{G}A\mathcal{F}p$  means that from every possible reachable state every way to continue we will eventually reach a state where the fact  $p$  is true.

We now make this intuition formal. As before, the definition builds truth values according to the hierarchical construction of the formula. Every state formula defines a set of states in which it is true. Path formulas are defined just like for LTL except that for the set of paths that start at a given state. Similar to the case of LTL, we denote by  $\mathcal{T}, s \models \varphi$  if a formula is satisfied in state  $s$  of  $\mathcal{T}$  and  $\mathcal{T}, s \not\models \varphi$  otherwise.

- If  $\varphi$  is a proposition, then  $\mathcal{T}, s \models \varphi$  if  $\varphi \in L(s)$  and  $\mathcal{T}, s \not\models \varphi$  otherwise.
- If  $\varphi$  is  $\neg\psi$  then  $\mathcal{T}, s \models \varphi$  if  $\mathcal{T}, s \not\models \psi$  and  $\mathcal{T}, s \not\models \varphi$  if  $\mathcal{T}, s \models \varphi$ . The definitions for formulas of the form  $\psi_1 \wedge \psi_2$  and  $\psi_1 \vee \psi_2$  is similar.
- If  $\varphi$  is  $E\mathcal{X}\psi$  then  $\mathcal{T}, s \models \varphi$  if there is a successor  $t$  of  $s$  (i.e.,  $(s, t) \in T$ ) such that  $\mathcal{T}, t \models \psi$  and  $\mathcal{T}, s \not\models \varphi$  if for all successors  $t$  of  $s$  we have  $\mathcal{T}, t \not\models \psi$ .
- If  $\varphi$  is  $E(\psi_1 \mathcal{U} \psi_2)$  then  $\mathcal{T}, s \models \varphi$  if there is some path  $\pi = s_0, s_1, \dots$  starting in  $s$  such that for some  $i$  we have  $\mathcal{T}, s_i \models \psi_2$  and for every  $0 \leq j < i$  we have  $\mathcal{T}, s_j \models \psi_1$  and  $\mathcal{T}, s \not\models \varphi$  if for every path  $\pi = s_0, s_1, \dots$  starting in  $s$  and for every  $i \geq 0$  if  $\mathcal{T}, s_i \models \psi_2$  then there is  $0 \leq j < i$  such that  $\mathcal{T}, s_j \not\models \psi_1$ .
- The definitions of other temporal connectives can be completed in a similar way by combining the path quantification with the definition from LTL.

We say that the transition system  $\mathcal{T}$  satisfies a CTL state formula  $\varphi$ , denoted  $\mathcal{T} \models \varphi$  if for every initial state  $s_0$  we have  $\mathcal{T}, s_0 \models \varphi$ . Otherwise,  $\mathcal{T}$  does not satisfy  $\varphi$ , denoted  $\mathcal{T} \not\models \varphi$ .

### 10.2.3 Model Checking

Model checking is the process of checking whether a formula holds for a specific transition system. That is, given a transition system  $\mathcal{T}$  and a formula  $\varphi$  (either in CTL or LTL), to decide if the formula is satisfied by the system, i.e., whether  $\mathcal{T} \models \varphi$ . In the case of LTL if the answer is negative we would like to get a run of the system that exhibits the failure of the property. That is, produce an initialized run  $r$  such that  $r, 0 \not\models \varphi$ .

The algorithmic approach towards model checking reduces this problem to a graph exploration problem. Essentially, we look on the transition system as a graph (sometimes with additional information) and deduce from analysis of this graph the correctness of the property. For both LTL and CTL we augment the transition system with auxiliary information. In the case of CTL the auxiliary information is by adding additional labels that tell us the truth values of simpler state formulas. In the case of LTL the algorithm is more complicated and the auxiliary information requires the addition of extra states and conditions. Here we give a short exposition of the

addition of labels that supports CTL model checking. A detailed exposition of LTL model checking is available for example in [2].

The algorithm for model checking CTL involves adding additional labels to the states of the transition system. Starting with a transition system  $\mathcal{T} = (S, T, S_0, \mathcal{P}, L)$  and a CTL property  $\varphi$  we show how to add labels to  $\mathcal{T}$ . We assume that  $\mathcal{T}$  already includes labels that tell us for every subformula of  $\varphi$  in which states it holds. Then, if  $\varphi$  is a Boolean combination of simpler operands then it is simple to deduce the truth of  $\varphi$  from the labels on states of  $\mathcal{T}$ . We simply add the label of  $\varphi$  to the states of  $\mathcal{T}$ . If  $\varphi$  is  $A\mathcal{X}\psi$ , then by assumption we have already included labels telling us when  $\psi$  is true in states of  $\mathcal{T}$ . It now suffices to iterate over all states of  $\mathcal{T}$ . If a state has a successor not marked by  $\psi$  then we do not change the label of such a state. If a state has all its successors marked by  $\psi$  then we add  $\varphi$  to the label of that state. The treatment of  $E\mathcal{X}\psi$  is similar.

We now turn to the treatment of  $E(\psi_1 \mathcal{U} \psi_2)$ . As before we assume that the labeling of states includes the value of  $\psi_1$  and  $\psi_2$ . We start by marking with  $E(\psi_1 \mathcal{U} \psi_2)$  all states that are marked by  $\psi_2$ . We then remove from our consideration states that are not labeled by  $\psi_1$ . Indeed, such states cannot be labeled by  $E(\psi_1 \mathcal{U} \psi_2)$ . We then repeatedly go over all states and add the marking  $E(\psi_1 \mathcal{U} \psi_2)$  to every state that has some successor marked by  $E(\psi_1 \mathcal{U} \psi_2)$  (considering only states marked by  $\psi_1$ ). Once no such states can be found this stage terminates. The description above sounds inefficient requiring repeated iteration over all the states. However, it can be implemented efficiently by iterating at most once over all possible transitions.

The treatment of  $A(\psi_1 \mathcal{U} \psi_2)$  is very similar. As before, we mark with  $A(\psi_1 \mathcal{U} \psi_2)$  all states that are marked by  $\psi_2$  and remove from consideration all states not marked by  $\psi_1$ . Then, the repeated iteration adds the label  $A(\psi_1 \mathcal{U} \psi_2)$  to states that have *all* their successors marked by  $A(\psi_1 \mathcal{U} \psi_2)$ .

The treatment of  $A\mathcal{F}\psi$  and  $E\mathcal{F}\psi$  is a special case of the two cases above, where we note that we can ignore the role of  $\psi_1$ . The treatment of  $E\mathcal{G}\psi$  and  $A\mathcal{G}\psi$  is by considering the equivalences  $E\mathcal{G}\psi \equiv \neg A\mathcal{F}\neg\psi$  and  $A\mathcal{G}\psi \equiv \neg E\mathcal{F}\neg\psi$ . We treat states not labeled by  $\psi$  as labeled by  $\neg\psi$ .

The algorithm for model checking LTL formulas is similar to the above in the sense that we create an algorithm that searches for paths in the graph obtained from the transition system and an additional structure obtained from the LTL formula.

### 10.3 Usage of Model Checking in Biology

In this section we survey a few modeling efforts that used model checking for analysis. We choose to highlight results that use the techniques as described above and that, in our opinion, show how model checking can be beneficial for the analysis of biological models. In some cases, these studies also led to the discovery of new biological insights. Obviously, in such a short book chapter it is impossible to survey all studies using model checking and we apologize to colleagues whose work we could not review due to lack of space.

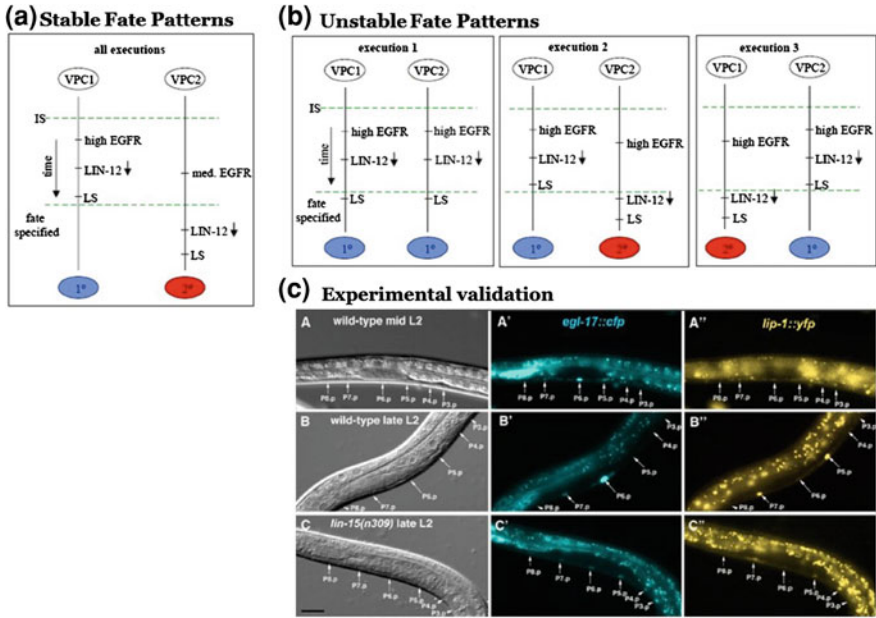
### ***10.3.1 Insights into Temporal Aspects of Signalling Crosstalk During Cell Fate Determination***

Describing mechanistic models in biology in a dynamic and executable language offers great advantages for representing time and parallelism, which are important features of biological behaviours. Model checking can be used to ensure the consistency of computational models with biological data on which they are based. Fisher et al. [20] have previously developed a dynamic computational model describing the mechanistic understanding of cell fate determination during the earthworm *C. elegans* vulval development, which provides an important paradigm for studying animal development. Model checking analysis of this model has provided new insights into the temporal aspects of the cell fate patterning process and predicted new modes of interaction between the signalling pathways involved. These biological insights, which were also validated experimentally, further substantiate the usefulness of dynamic computational models to investigate complex biological behaviours.

Fisher et al. [20] have used model checking for two purposes. First, to ascertain that their mechanistic model reproduces the biological behaviour observed in different mutant backgrounds. For that, they have formalized the experimental results described in a set of papers and verified that all possible executions satisfy these behaviours. That is, regardless of the order of interactions from a given set of initial conditions, different executions always reproduced the experimental observations. Second, they also used model checking to query the behaviour of the model. By phrasing queries such as which mutations may lead to a stable or an unstable fate pattern, they analyzed the behaviour of the model. Once an unstable mutation was found, they determined what part of the execution allows this kind of mutations by disallowing different behavioural features of the model and checking when the instability disappears (see Fig. 10.2).

By using model checking to compare the executable model with existing experimental data, Fisher et al. predicted novel interactions in the signalling network governing vulval fate specification, in addition to predicting the outcome of perturbations that are difficult to test experimentally. These insights led to suggest a revised model with at least one additional negative feedback loop indicated by the inhibition arrow in Fig. 10.3.

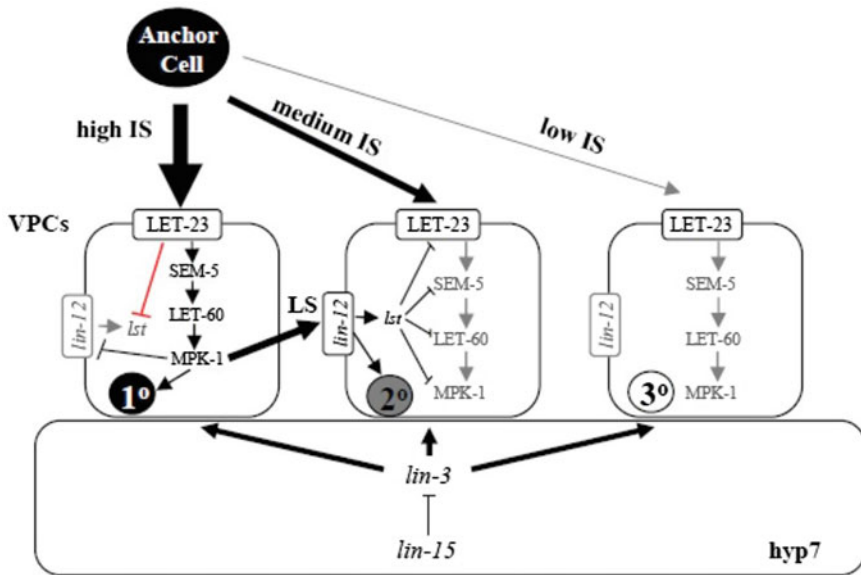
Through model checking an executable model representing the crosstalk between Epidermal growth factor receptor (EGFR) and LIN-12/Notch signalling during *C. elegans* vulval development Fisher et al. gained new insights into the usage of these conserved signalling pathways that control many diverse processes in all animals. While many modelling efforts use simulations that allow investigating only a few possible executions, this work had emphasized the power of analyzing all possible executions using model checking.



**Fig. 10.2** Order of events in stable and unstable fate patterns and experimental validation. The *upper panels* show sequence diagrams. Time flows from *top to bottom*. Two events that appear on the same *vertical line* are ordered according to the time flow. The *dashed lines* synchronize the different *vertical lines*. All events that appear above a synchronization line occur before all events that appear below the synchronization line. The time-order between two events that appear on *parallel vertical lines* without a synchronization line is unknown. **a** Proposed sequence of events leading to a stable pattern. The left time line starts with a high inductive signal (*IS*) and the right time line (**b**) with a medium *IS*. **b** Three diagrams that represent possible sequences of events leading to different fate patterns. **c** Experimental validation of the loss of sequential activation in *lin-15* mutants, as predicted by the computational model. The pictures visualize cell fate specification in (**c**). *Elegans* with *blue* and *yellow* fluorescent proteins (*EGL-17::CFP* and *LIP-1::YFP*) expressed during activation of the inductive and lateral signaling pathways, respectively. The *upper and middle rows* show examples of wild-type animals at mid and late L2 stage expressing the *EGL-17* marker in *P6.p* and the *LIP-1* marker in *P5.p* and *P7.p*, respectively. The *lower row* shows examples of a *lin-15* mutant at the late L2 stage showing simultaneous expression of both markers in *P5.p* and *P7.p*. Reproduced from [20]

### 10.3.2 Symbolic Analysis of Biochemical Networks

The idea to use computation tree logic (CTL) as a query language for biochemical networks was first introduced by Fages, Schächter and colleagues in 2004. Chabrier-Rivier et al. were the first to show the potential of using symbolic model-checking tools to evaluate CTL queries in the context of mammalian cell-cycle control and gene expression regulation [11, 12]. More recently, the Biochemical Abstract Machine (BIOCHAM) tool was introduced as an environment to model and analyse biochemical networks using model checking [10]. BIOCHAM is based on a rule-based

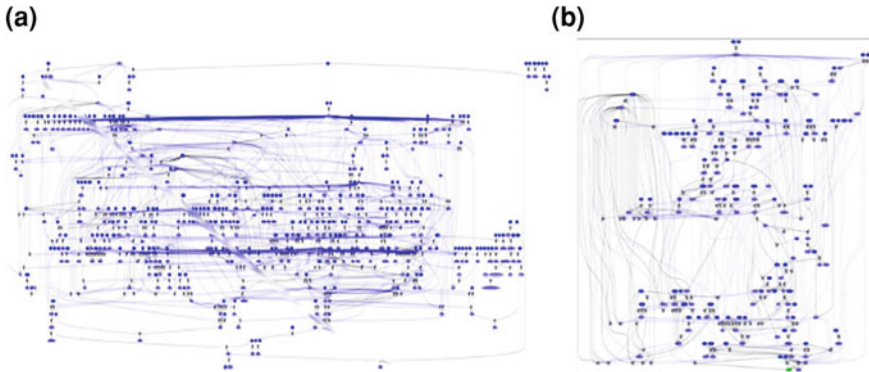


**Fig. 10.3** Conceptual model for the signaling events underlying VPC fate specification. Diagrammatic mechanistic model for the signaling events underlying vulval precursor cell (VPC) fate specification. Inductive signal (*IS*), lateral signal (*LS*), cell fates: primary 1°, secondary 2°, tertiary 3°. Reproduced from [20]

language that allows the user to write models of biochemical networks and perform multi-level analysis, and a temporal logic language (CTL or LTL) used to formalize experimental data. BIOCHAM permits continuous or stochastic simulations, as well as model validation or revision according to a formal qualitative or quantitative specification. Consequently, BIOCHAM allows to verify that whenever an interaction or a molecule is added to the network, the global properties of the system (expressed in temporal logic) are conserved. In addition, it is possible to automatically search for parameter values that reproduce the specified behaviour of the system under different conditions.

The *Pathway Logic* [18] is tightly related to the symbolic model checking approach as it is based on rewriting logic in order to model and analyze signal transduction and biochemical networks, and interpret experimental data. In *Pathway Logic*, biological molecules, their states, locations, and their role in molecular or cellular processes can be modelled at different levels of abstraction. An example of an EGFR pathway model as a Petri net is shown in Fig. 10.4a. *Pathway Logic* can be used to browse the model and ask for subnets or pathways satisfying goals of interest (Fig. 10.4b) [33].

The modelling of biochemical networks with concurrent transition systems is of a somewhat lower level than with *Pathway Logic*. *Pathway Logic* is more expressive as it can express algebraic properties of the components, such as the associativity of



**Fig. 10.4** Model of EGF stimulation using *Pathway Logic*. **a** An impression of the Pathway Logic Assistant (*PLA*) rendering of the model as a Petri net. **b** The subnet of all reactions relevant to activation of Erk in response to a stimulus by EGF is obtained by making Erk1 (and/or Erk2) a goal and asking *PLA* for the subnet. Reproduced from [33]

complexation. This capability can be used to infer the possible reactions of molecules from their logical structure.

### 10.3.3 Insights into Signalling Crosstalk During Pancreatic Cancer

Clarke and colleagues applied symbolic model checking to study temporal logic properties in a model of pancreatic cancer. This is the first in-silico model describing the crosstalk between six signalling pathways that have genetic alterations in all pancreatic cancers, with the aim to investigate apoptosis (programmed cell death), proliferation, and cell-cycle arrest. The signalling network model composed of the EGF-PI3K-P53, Insulin/IGF-KRAS-ERK, SHH-GLI, HMGB1-NFkB, RBE2F, WNT-b-Catenin, Notch, TGF b-SMAD, and apoptosis pathway verified temporal logic properties encoding behaviour related to cell fate, cell cycle, and oscillation of expression level in key proteins. The model agreed well with experimental observations as well as suggested several properties to be tested by experimentally.

## 10.4 Underlying Techniques

In this section we revisit the algorithms for model checking and expose some of the techniques used to implement those algorithms efficiently. In general, we term these techniques as *graph representation* and *graph analysis* techniques. After exposition of these techniques we survey some results where these techniques were used for



analysis of biological models. In these cases, the knowledge of the techniques used by model-checking tools for the effective analysis of transition systems were used to effectively analyze biological models.

### 10.4.1 Symbolic Transition Systems

In Sect. 10.2 we concentrated on the explicit representation of models. That is, every snapshot of the status of the system was treated individually. While this approach is very intuitive it has its limitations. Most importantly the size of models that can be handled. For example, a moderately sized model that represents the status of 30 substances, each represented as either active or inactive, has  $2^{30}$  states, which is about one Billion. Approaches that call for direct drawing of the possible transitions of such a model for user inspection are hopeless. But even exploring each one of these states automatically will incur a significant time delay. Alternative approaches to represent the states and executions of systems have been developed [7–9]. Such approaches are generally termed *symbolic* and they raise the level of representation from that of single states to that of sets of states.

We explore an alternative *symbolic* representation of a transition system. A symbolic transition system is  $\mathcal{T} = (V, \rho, \Theta)$  with the following components:

- $V$  is a set of variables, each ranging over a fixed range  $R(V)$ . By writing formulas over the variables in  $V$  we can represent sets of states. For example, the formula  $v_1 > 2 \wedge v_2 \leq 3$  represents all the states where the value of  $v_1$  is more than 2 and  $v_2$  is at most 3. A *valuation*  $\sigma$  of the system is an assignment of value to each variable  $v \in V$  such that  $\sigma(v) \in R(v)$ . The language used for representing such formulas depends on the ranges of variables. Here, we assume that all variables range over the natural numbers or (small) sets of natural numbers. We restrict attention to formulas constructed by taking the normal arithmetic operations over natural numbers and variables, comparison between such expressions, and usage of conjunction, disjunction, and negation to combine such terms to large formulas. Thus, formulas can represent the set of valuations such that by assigning the value of the variables into the formula it evaluates to true. In order to represent transitions we use two copies of the variables. We take a disjoint copy of the variables  $V$  and add primes to them  $V'$ . The primed variables represent the *next* values of the variables. Thus, by writing a formula like  $v_1 = 1 \rightarrow v'_1 = 2$  we impose the restriction that whenever  $v_1$  is 1 it changes its value to 2. By writing formulas over  $V \cup V'$  we can represent changes in the values of variables, or transitions between states of the model. Thus, a pair of valuations  $\sigma$  and  $\sigma'$  satisfies a formula over  $V \cup V'$  if by assigning the values in  $\sigma$  to all the variables in  $V$  and the values in  $\sigma'$  to all the variables in  $V'$  the formula evaluates to true.
- Accordingly,  $\Theta$  is a formula over the variables  $V$  representing the initial states and  $\rho$  is a formula over  $V \cup V'$  representing the transitions.



It is simple to convert the symbolic representation to the explicit representation presented in Sect. 10.2. Given a symbolic transition system  $\mathcal{T} = (V, \rho, \Theta)$  and a set of formulas  $\varphi_1, \dots, \varphi_n$  representing the basic propositions about the model we can construct the explicit transition system  $\mathcal{T} = (S, T, S_0, \mathcal{P}, L)$ , where  $S$  is the set of possible assignments to variables in  $V$ , i.e., all the valuations of  $\mathcal{T}$ ,  $S_0$  is the set of states/valuations satisfying the formula  $\Theta$ ,  $T$  is the set of pairs of states/valuations satisfying the formula  $\rho$ . Finally, the propositions  $\mathcal{P}$  are  $\{\varphi_1, \dots, \varphi_n\}$  and  $L$  associates with a valuation  $\sigma$  all the propositions that hold true in that valuation.

It seems that we have made a meaningless change of notation. However, the truth could not be further. Now, we have a way of easily representing sets of states as formulas. Furthermore, by manipulating such formulas we can manipulate sets of states. For example, taking two sets of states represented by formulas  $\phi_1$  and  $\phi_2$  the formula representing their intersection is  $\phi_1 \wedge \phi_2$  and the formula representing their union is  $\phi_1 \vee \phi_2$ . We can check set equivalence by testing formula equivalence and check whether the set of states represented by a formula is empty by checking whether the formula is satisfiable. By introducing quantification over variables we also can compute the set of successor states of a given set. That is, if  $\phi$  represents a set of states and  $\rho$  is the transition relation over variables  $V \cup V'$  then the following formula represent the set of states that can be reached from  $\phi$  in one application of  $\rho$ :

$$unprime(\exists V.\phi \wedge \rho)$$

That is, the formula  $\phi \wedge \rho$  represents the pairs of valuations  $(\sigma, \sigma')$  such that  $\sigma$  satisfies  $\phi$  and  $\sigma'$  is a successor of  $\sigma$  as  $(\sigma, \sigma')$  satisfies  $\rho$ . Then,  $\exists V.\phi \wedge \rho$  throws away the variables in  $V$  leaving a formula over  $V'$  that characterizes valuations over  $V'$  such that there is some value for the variables in  $V$  such that  $\phi \wedge \rho$  holds for the pair. Exactly the states that are successors of states in  $\phi$ . Finally, we have to translate every reference in  $\exists V.\phi \wedge \rho$  to a variable in  $V'$  to refer to the same variable in  $V$ . This is the role of the *unprime*. It follows that we have a symbolic way to represent the application of the transition to a set of states. We denote this in short as  $next_\rho(\phi)$ . Similarly,  $prev_\rho(\phi) \equiv \exists V'.(\rho \wedge prime(\phi))$ , computes the set of predecessors of  $\phi$ . The *prime* operator changes references to  $V$  to references to  $V'$  resulting in a formula that characterizes all the valuations over  $V'$  that satisfy  $\phi$ . Then, adding  $\rho$  ensures that we characterize pairs satisfying the transition such that the second satisfies the formula  $\phi$ . Throwing away the variables in  $V'$  we get the desired formula.

Now we need two additional tools. Suppose that all variables in  $V$  range over finite domains. Then, a variable  $v$  ranging over  $\{1, \dots, n\}$  can be represented by  $\log(n)$  Boolean variables. It follows that we can think about formulas over Boolean variables and in order to use the symbolic approach we need an efficient way to store, manipulate, and evaluate formulas over Boolean variables. Boolean Decision Diagrams (BDDs for short) [8] do exactly that. They are a canonical representation of Boolean formulas making comparison between such formulas very simple. Operations such as conjunction, disjunction, and negation can be implemented efficiently. Finally, existential quantification is done by translating it to disjunction.

The second tool is satisfiability solving. If variables range over finite domains we can translate them to Boolean variables as above and we need a solver for propositional formulas [7]. If variables range over infinite domains we need a theorem prover [31] or an SMT solver [23]. We are now in position to use the symbolic representation for analysis.

### 10.4.2 Symbolic Model Checking

The algorithms in Sect. 10.2 consisted of annotating states by additional markings corresponding to the CTL formulas holding in them. Here we use formulas to represent the same.

Suppose that we have computed a formula representing the set of states satisfying a CTL formula  $\psi$ . This is straight forward for propositions as they are already formulas representing sets of states. Consider a formula of the form  $\varphi = E\mathcal{X}\psi$ . Then,  $prev(\psi)$  is the formula representing the set of states satisfying  $\varphi$ . The set of states satisfying a formula of the form  $\varphi = A\mathcal{X}\psi$  is represented by  $\neg prev(\neg\psi)$ . We have already explained how to handle Boolean connectives and we turn now to the until operator. For a formula  $E(\psi_1\mathcal{U}\psi_2)$  we do the following inductive process. We start with  $\phi_0 = \psi_2$ , and then compute  $\phi_{i+1} = \phi_i \vee (\psi_1 \wedge prev(\phi_i))$ . We then compare  $\phi_{i+1}$  to  $\phi_i$ . If they are equivalent, the process has terminated and we have computed the formula representing the set of states satisfying  $E(\psi_1\mathcal{U}\psi_2)$ , otherwise, we proceed with another step. Similarly,  $A(\psi_1\mathcal{U}\psi_2)$  is computed by iterating over  $\phi_0 = \psi_2$  and  $\phi_{i+1} = \phi_i \vee (\psi_1 \wedge \neg prev(\neg\phi_i) \wedge prev(true))$ . The need in adding  $prev(true)$  is to avoid adding states satisfying  $\psi_1$  that have no successors at all, which obviously do not satisfy  $A(\psi_1\mathcal{U}\psi_2)$ .

It turns out that in practice, symbolic model checking, in many cases, outperforms explicit model checking. These techniques combined with BDD representations allowed model checking of hardware designs to scale to systems composed of  $10^{120}$  states and more [9].

### 10.4.3 Path Representation

In recent years efficient satisfiability solvers and SMT solvers have been developed [17, 27, 28]. These tools enabled a different approach to model checking. This approach creates a formula representing a set of executions of the model. By asking whether this formula is satisfiable we can search for paths of a certain length. Furthermore, by combining the formula representing executions with additional constraints on the states participating in such executions we can search for executions satisfying given conditions. For example, one could be looking for executions reaching a certain state or set of states. Alternatively, one could search for paths satisfying a sequence of conditions or evolving according to a certain pattern.

Consider a system  $\mathcal{T} = (V, \rho, \Theta)$ . In order to represent executions of  $\mathcal{T}$  of length  $n$  we create  $n$  copies of  $V$ . That is,  $V_0, \dots, V_{n-1}$  are all copies of the variables in  $V$  each numbered with the location in the execution. A valuation to the variables in  $V_0, \dots, V_{n-1}$  is now a representation of  $n$  states of the system. We now create a formula representing executions of length  $n$  by translating  $\Theta$  to  $\Theta_0$ , which refers to the first copy  $V_0$  instead of  $V$  and translating  $\rho$  to  $\rho_{i,j}$ , which refers to  $V_i$  instead of  $V$  and to  $V_j$  instead of  $V'$ . Thus, the formula  $P_i \equiv \Theta_0 \wedge \bigwedge_{i=0}^{n-2} \rho_{i,i+1}$  is a formula over the variables  $V_0, \dots, V_{n-1}$ . A satisfying assignment to the  $P_i$  is a sequence of states such that the first satisfies  $\Theta$  (through  $\Theta_0$ ) and every pair of adjacent states satisfies  $\rho$  (through  $\rho_{i,i+1}$ ). The formula  $L_i \equiv \Theta_0 \wedge \bigwedge_{i=0}^{n-2} \rho_{i,i+1} \wedge \bigvee_{i=0}^{n-1} \rho_{n-1,i}$  is satisfiable if there exists a looping execution of length at most  $n$ .

#### 10.4.4 Biological Model Analysis

We now survey results that take advantage of the techniques underlying model checking to improve analysis of biological models. Both apply to the analysis of discrete models that extend Boolean networks; Qualitative Networks (QNs, for short) [32] and Gene Regulatory Networks (GRNs, for short) [34]. We give a short informal introduction to these formalisms and explain how model checking techniques are used to analyze them.

We give a short exposition of QNs and how they give rise to transition systems. A model in *Qualitative Networks* includes variables that represent the concentration of proteins as a discrete value in a (small) fixed range. Changes in variable values are gradual allowing them to increase or decrease by at most 1 in every step of the system. Mathematically, a QN  $Q$  includes two components  $(V, T)$ : The set  $V = \{v_1, \dots, v_n\}$  is a set of variables each ranging over a finite range  $D(v_i) = \{j, \dots, k\} \subseteq \mathbb{N}$ , e.g.,  $\{0, 1, 2, 3\}$ . The set  $T = \{T_1, \dots, T_n\}$  include target functions for all variables. A *state* of  $Q$  is an assignment  $s : V \rightarrow \mathbb{N}$  such that for every  $i$  we have  $s(v_i) \in D(v_i)$ . Let  $S$  denote the set of all possible states of the QN. Each *target function*  $T_i \in T$  associates the target value of variable  $v_i$  for every state of the system. Formally,  $T_i : S \rightarrow D(v_i)$ . Intuitively, in state  $s \in S$ , variable  $v_i$  will change to get to the value  $T_i(s)$ , however will do so in increments/decrements of 1. It follows that a QN  $Q$  gives rise to a transition system  $\mathcal{T}_Q = (S, \Delta, S, \mathcal{P}, L)$ , where the components of  $\mathcal{T}_Q$  are as follows.

- $S$  is the set of states as explained above and all states are initial.
- $\Delta$  is the set of transitions that associates with state  $s$  the successor  $s'$  such that for every variable  $v_i \in V$  we have

$$s'(v_i) = \begin{cases} s(v_i) + 1 & \text{If } T(s) > s(v_i) \quad \text{and } s(v_i) + 1 \in D(v_i) \\ s(v_i) - 1 & \text{If } T(s) < s(v_i) \quad \text{and } s(v_i) - 1 \in D(v_i) \\ s(v_i) & \text{Otherwise} \end{cases}$$

- The set of propositions  $\mathcal{P}$  is  $v_i = j$  for  $v_i \in V$  and  $j \in D(v_i)$ .
- The labeling  $L$  associates with a state  $s$  the propositions  $v_i = s(v_i)$  for every  $v_i \in V$ .

Thus, the transition system updates all the variables in the system by allowing them to pursue their target by a change of at most 1. The basic facts labeling each state of the model are the values of the different variables.

One of the most interesting questions regarding qualitative networks has been that of *stabilization*. A network is called stabilizing if there is a unique state  $s$  such that  $\Delta(s, s)$  and for every other state  $t$  it is impossible to get from  $t$  to itself by application of  $\Delta$ . That is, for every  $t_1, \dots, t_n$  such that  $\Delta(t_i, t_{i+1})$  for every  $1 \leq i < n$  we have  $t_1 \neq t_n$ . The question of stabilization can be answered by computing the set of states that are visited along arbitrarily long paths. If that set has the size 1, then the network is stabilizing. This observation suggests the following algorithm for checking stabilization. Let  $R_0 = S$  and for  $i \geq 0$  let  $R_{i+1} = \Delta(R_i) = \{s \mid \exists t \in R_i \text{ s.t. } \Delta(t, s)\}$ . It is clear that  $R_{i+1} \subseteq R_i$ . It follows that if a state appears on a cycle in  $\mathcal{T}_Q$ , it remains in  $R_i$  for all  $i$ . It is equally easy to see that if a state  $s$  is not on a cycle in  $\mathcal{T}_Q$  then for some  $i$  it does not appear in  $R_i$ . Then, by repeatedly computing  $R_i$  for increasing values of  $i$  one can find a set  $R_i$  such that  $R_i = R_{i+1}$ . It follows that this set  $R_i$  includes exactly the set of states that appear on cycles in  $\mathcal{T}_Q$ . Unfortunately, the straight forward computation of  $R_i$  has been elusive. In [32], it was suggested to iteratively compute  $R_i$  by abstracting parts of the system. This abstraction lead to a considerable increase in capacity of networks that can be analyzed compared with the naïve approach. In [16], it was suggested that instead of constructing an exact representation of the set  $R_i$ , it would be enough to consider subsets of  $R_i$  for which the range of each variable is a contiguous range of values. For example, in a system with two variables  $v_1$  and  $v_2$  ranging over  $\{0, \dots, 4\}$  the set that includes the two points  $(0, 1)$  and  $(1, 0)$  would be represented by the set of states in which  $0 \leq v_1 \leq 1$  and  $0 \leq v_2 \leq 1$ . Then, the set  $R_{i+1}$  is the best set of this form that captures the transition of  $Q$ . The results in [16] show that this abstraction technique scales to an order of magnitude larger models than those that could previously be handled. This technique has been made available through the tool BMA [5].

We now turn to survey the usage of model checking in the analysis of GRNs. The transition structure of GRNs and QNs are very similar. The main difference is that in the context of GRNs the target functions are defined in terms of so called *parameters*. Somewhat simplifying presentation, the parameters are the actual value of the target function per each possible value of the inputs. For example, consider a system with variables  $v_1, v_2$ , and  $v_3$  all ranging over  $\{0, 1, 2\}$ . If  $v_1$  is affected by  $v_2$  and  $v_3$  then the parameters are essentially the values written in the  $3 \times 3$  table for all possible options for the values of  $v_2$  and  $v_3$ . An entry in the table is the value of the target function of  $v_1$  when  $v_2$  and  $v_3$  are in the appropriate values matching this table entry. One of the interesting usages of model checking in the context of GRNs is to try to narrow down the space of possible values of parameters. In this context, a GRN is given without concrete knowledge of the parameters but with additional dynamic behaviors that are exhibited in experimental results. These dynamic behaviors should be translated

to temporal logic specifications that talk about the changes in variable values over time. Then, model checking can be used to check which parameter values allow the network to reproduce this behavior. We note that in this context, model checking is used dually to the way it was described above. Thus, instead of requiring that all behaviors for a given parameter value reproduce the specification we require that for an *admissible* parameter value there be a behavior reproducing the specification.

Technically, in order to enable this check, one has to augment the transition system with variables that represent the parameter values. The transition system encodes the fact that parameter values do not change over an execution and that the changes in the values of “core” GRN variables depend on the values of the parameter variables. Then, a model checking query can create a representation of the set of possible parameter values that allow a particular behavior. This set can be narrowed down by considering multiple dynamic behaviors. See, for example, recent work on such usage of Model Checking in the context of GRNs [3, 4, 6].

## 10.5 Probabilistic Model Checking

So far the formalism we used to represent models has been transition systems. Transition systems can represent possible transitions and possible executions, however, they have no quantitative information regarding the prevalence of these transitions. In order to include such information in our models we have to use richer formalisms that include such information. In this section we give a short exposition of continuous time Markov chains (CTMCs). These are models that have discrete states and cannot represent continuous changes in values of variables. However, they do capture probabilities in change from state to state and include also a representation of continuous time. The algorithms involved in the analysis of Markov chains are significantly more complex and we do not review them here. The interested reader is referred to [2, 25, 26].

### 10.5.1 Markov Chains and Their Analysis

We extend the concept of a transition system to that of a continuous-time Markov chain (CTMC). For that, we replace the transitions with probabilistic transitions. CTMCs are usually defined via higher level languages that resemble the symbolic representation of transition system. We try to keep the discussion as general as possible and avoid relating to such formalisms. We then introduce the temporal logic continuous stochastic logic (CSL) that is used to express properties of CTMCs.

A CTMC is  $\mathcal{T} = (S, T, s_0, \mathcal{P}, L)$ , where  $S$  and  $L$  are as in transition systems. The set of transitions  $T: S \times S \rightarrow \mathbb{R}^+$  associates with every pair of states  $(s, t)$  a non-negative real number  $T(s, t)$  corresponding to the *rate* of transition from state  $s$  to state  $t$ . There is one initial state  $s_0 \in S$ . The transition  $T(s, t) = r > 0$  implies that

it is possible to change from state  $s$  to state  $t$ . The change occurs within  $t$  time units with probability  $1 - e^{-r \cdot t}$ . If multiple transitions from  $s$  are possible, i.e.,  $T(s, t) > 0$  and  $R(s, t') > 0$  for  $t \neq t'$ , then there is a *race* between the transition to  $t$  and the transition to  $t'$ . The *exit rate* from state  $s$  is the sum of positive rates leaving state  $s$ , that is  $T(s) = \sum_t T(s, t)$ . The probability that some transition occurs within  $t$  time units is  $1 - e^{-T(s) \cdot t}$  and the probability to end up in state  $t''$  is  $R(s, t'')/T(s)$ . One interesting features of probabilistic systems is that an experiment that has positive probability and is tried often enough will eventually succeed. Accordingly, if a state of the CTMC is revisited often enough every one of its successors will be visited often enough. This leads to the steady state behavior of a CTMC. In the long run, some of the states of the CTMC will be transient: the amount of time spent in them will be 0, and some recurrent: the amount of time spent in them will be greater than 0. Accordingly, the value  $R_\infty(s, t)$  is the probability of having started in state  $s$  to be in state  $t$  in the long run.

We are now ready to define CSL, a version of temporal logic that is used to specify properties of CTMCs [1]. CSL extends CTL by adjusting it to continuous time and replacing path quantification by probabilistic quantification. As before CSL combines the basic facts with Boolean operators and special temporal operators.

- The temporal operators  $\mathcal{X}$  and  $\mathcal{U}$  are nested within probabilistic path quantification  $[\cdot]_{\bowtie p}$ , where  $\bowtie \in \{>, \geq, <, \leq\}$  and  $p \in [0, 1]$ . Furthermore  $\mathcal{U}$  is extended by allowing to state an interval  $I$  in which the second formula is expected to happen. Thus,  $[\mathcal{X}\phi]_{>0.5}$  means that the measure of paths that satisfy  $\phi$  in the next state is more than 0.5. Similarly,  $[\phi_1 \mathcal{U}^{[5,7]} \phi_2]_{\leq 0.2}$  means that the measure of paths that satisfy that  $\phi_1$  holds until  $\phi_2$  holds sometime between 5 and 7 time units from now is at most 0.2.
- The steady-state operator  $S(\cdot)_{\bowtie p}$  states that in the long run the probability to be in states that satisfy the subformula must satisfy the probability condition.

As before we make this intuition more formal. Every formula defines a set of states in which it is true. Just like CTL and LTL we denote by  $\mathcal{T}$ ,  $s \models \varphi$  if formula  $\varphi$  is satisfied in state  $s$  of  $\mathcal{T}$  and  $\mathcal{T}, s \not\models \varphi$  otherwise.

- For propositions and Boolean operators the definitions is just as before.
- If  $\varphi$  is  $[X\psi]_{\bowtie p}$  then  $\mathcal{T}, s \models \varphi$  if  $prob_{\mathcal{T}}(\pi, \varphi) \bowtie p$ . That is, if the probability of the set of paths that start in  $s$  and in the next state visit states that satisfy  $\psi$  satisfies the comparison with  $p$ .
- If  $\varphi$  is  $[\psi_1 \mathcal{U} \psi_2]_{\bowtie p}$  then  $\mathcal{T}, s \models \varphi$  if  $prob_{\mathcal{T}}(\pi, \varphi) \bowtie p$ . That is, if the probability of the set of paths that start in  $s$  and satisfy  $\psi_1 \mathcal{U} \psi_2$  such that  $\psi_2$  is visited in the interval  $I$  satisfies the comparison with  $p$ .
- If  $\varphi$  is  $S[\psi]_{\bowtie p}$  then  $\mathcal{T}, s \models \varphi$  if  $\sum_{\mathcal{T}, s' \models \psi} R_\infty(s, s') \bowtie p$ . That is, if starting in  $s$  the long term probability of visiting states that satisfy  $\psi$  satisfies the comparison with  $p$ .

### 10.5.2 *Biological Modeling with Markov Chains*

Using CTMCs encoding of molecular networks is very direct. A typical model includes a certain set of species of molecules. A state of a model represents the number of molecules of each species (or the concentration level) and transitions correspond to interactions between substances [30].

More formally, a molecular model is defined over a set of substances  $V$ . For each substance  $v \in V$  one defines the number of levels  $l(v)$  of  $v$ , implicitly deciding whether  $v$  represents molecules explicitly or their concentration level (cf. [13, 22]). Furthermore, the model includes molecular interactions and their rates. For example, phosphorylation would correspond to one molecular species changing to another (non-activated to activated). Similarly, binding of two molecules would correspond to the number of the two simple forms decreasing and the number of the bound substance increasing. Unbinding would be treated dually. So in a state  $v$ , each molecular interaction would correspond to a transition that changes the numbers of substances in  $v$  according to one interaction taking place. The rate with which the interaction takes place would then depend on the number of possible molecular interactions. Thus, if an interaction is a change from one molecular type to another, its rate is a product of the rate of a single such interaction with the number of instances of the source type in  $v$ . If an interaction is a binding, its rate is a product of the number of possible pairs of molecules and thus is a product of the rate of a single such interactions with the number of instances of the first source and the number of instances of the second source. The disadvantage of this approach is that it leads to huge models with many possible transitions enabled from each state. This is especially true if molecules are modeled individually and not through concentration levels. This severely limits the size of models that can be analyzed by model checking. Still, the extensive analysis enabled by model checking makes it useful to analyze even models of bounded size.

One example of an application of such an analysis is the model by Heath et al. [22] of the Mitogen-activated protein kinase (MAPK). In this chapter a CTMC model of the Fibroblast growth factor (FGF) signalling pathway is constructed. Due to the scalability issue mentioned above the model is analyzed with very low number of copies of each molecules. The model is analyzed with questions such as: What is the probability that a certain species is bound to another species at a given time  $t$ ? What is the expected number of times that a molecule binds before degrading? The exact analysis of these questions sheds light on the roles of different molecules within the pathway. A similar (in terms of the technology involved) study of the MAPK pathway is available in [24].

Another example is the model of gp130/JAK/STAT signalling pathway using the concentration level approach [21]. Here, probabilistic model checking is used to ensure that the model is of sufficient quality. For example, model checking identifies that one of the substances can “run out” in the system and lead to no further molecular interactions being possible. This highlights the important role of this substance and the need in further modeling of the production of this substance. Then, analysis



similar to that described above shows that a full phosphorylation of some of the substances is achieved with high probability.

## 10.6 Lessons Learned

To summarize, we highlight the main issues that we believe are important for the development of this approach:

- Model checking is a powerful technique for the analysis of programs. Over the last decade model checking has been successfully used for the analysis of biological models, providing novel insights into various cellular mechanisms and behaviors.
- Many tools providing implementations of model checking are experimental and academic in nature. This implies that users require certain expertise in underlying techniques and formalisms in order to use model checking. The development of more reliable and user-friendly tools, as well as approaches that facilitate the creation of models, could further encourage users with little to no formal background to use these tools for biological modeling and analysis.
- Biological models are somewhat different from software and hardware programs. This calls for the formal methods community to develop dedicated techniques and algorithms that are particularly tailored for the analysis of biological models, leading to improved capacity and efficiency when analyzing biological models.
- Model checking cannot stand on its own as a sole technique of analysis. It is crucial to combine multiple forms of analysis of the same model. One of the major obstacles to combine multiple forms of analysis is the lack of standardization in modeling languages. While SBML provides a standard for mathematical biological models, a similar language that supports further types of models is missing. Such a language could provide a cross-tool foundation for sharing and distributing models enabling analysis by multiple approaches.

## 10.7 Glossary

- *Reactive Systems*: A system that consists of parallel processes, where each process may change state in reaction to another process changing state. Biological systems are highly reactive (e.g., cells constantly send and receive signals and operate under various conditions simultaneously).
- *Formal methods*: A collection of methods that relate to formal logic to analyze computer systems and prove properties (written in formal logic) about such systems.
- *Model checking*: A technique for proving that systems have certain properties described in *temporal logic*. In case of proof failure, in most cases, the technique provides a *counter example*—an execution that violates the requirement.



- *Nondeterministic system*: A system that may have several possible reactions to the same stimulus. In biological systems, for example, we can observe various patterns of cell fate under the same genotype. Hence, nondeterministic models capture the diverse behavior often observed in biological systems by allowing different choices of execution, without assigning priorities or probabilities to each choice.
- *Property/Requirement/Specification*: A formal sentence describing some aspect of a program or system.
- *Transition system*: A computational model for a system. A *state* of a transition system describes the status of the world (restricted to the point of interest of the model/system under study). *Transitions*, which are connections between states describe, the possible changes to the world.
- *Temporal logic*: A specific formalism for describing properties of systems. Temporal logic describes possible evolutions of systems over time. Generally classified to *linear time* or *branching time* according to their view of a computation. In the linear time view a computation is a sequence of the states of the system. Nondeterministic systems have multiple possible computations. In the branching time view a computation is a tree like structure encompassing all possible options of the system. A nondeterministic system has one computation that resembles a tree. *Linear temporal logic (LTL)* and *Computation Tree Logic (CTL)* are examples of a linear time and a branching time logic often used in verification of computer systems.
- *Logic operators*: The combinators of simple logic formulas to more complicated ones. For example, an “and” operator combines two Boolean operands. *Binary operators* and *unary operators* operate on two or one operands respectively. In the context of temporal logic we distinguish between *Boolean operators*, which combine the truth values of formulas at a given point in time (e.g., “and”, “or”, or “not”), and *Temporal operators*, which combine the truth values of formulas in different time points.
- *Boolean networks*: Computational models that describe a biological system by referring to its components as either “active” or “inactive”. Usually, each component relates to a certain protein. Components change their values according to positive and negative influences from other components.
- *Petri nets*: Computational models that describe the state of the world by associating a number of “tokens” with designated “places”. “Transitions” prescribe how tokens can move from place to place leading to a general change of conformation of the system.
- *Graph representation*: Representing a transition system in the form of a mathematical graph. Nodes of the graph correspond to the states of the system and (directed) edges of the graph correspond to transitions of the system.
- *Graph analysis*: Applying algorithms on the graph representation of the system.
- *Symbolic model checking*: Applying the model checking technique by combining reasoning over sets of states instead of reasoning over individual states. Using such techniques model checking can scale to systems with a huge number of states that cannot be enumerated.

- **prime operator:** In a symbolic representation of a transition system this is our way of saying that a variable relates to the next time point and not the current one.
- **Boolean Decision Diagrams (BDDs):** A specific technique for storing sets of states through relating to them as Boolean functions.
- **Satisfiability solver:** A tool that solves the question of whether a Boolean formula is satisfiable. A Boolean formula is a way of stating constraints over the values of Boolean variables. The formula is then satisfiable if there is a way to assign Boolean values to variables so that the formula evaluates to true.
- **SMT solver:** A tool that solves the question of whether a formula that combines Boolean parts and additional (theories) parts is satisfiable.
- **Qualitative networks:** An extension of Boolean Networks that allows more values to represent the possible status of each component and allows a more flexible way of describing how the values of components change over time. The changes in the values of variables are defined through so called target functions, which describe the value that the component aspires to get to. The value of the component then changes gradually until it attains this target.
- **Fixed point:** a value in a computation that does not change when applying to it some operation. This is used many times to describe states of a system that does not change anymore. Also, in algorithms that compute a set of states by applying a certain operation to them a fixpoint is a set of states that the operation does not change.
- **Stabilization:** One of the main properties checked for Boolean networks and Qualitative networks. Essentially, this is a property of the system which indicates that the system has exactly one fixpoint. That is, there is a unique stabilization state such that regardless of the starting values of the components in the network, after a long enough execution the stabilization state is reached and never changed anymore.
- **Continuous time Markov chains (CTMCs):** A computational model combining discrete state transitions with continuous time and probabilities. As in general transition systems, the state of the world is described via a “state”, however, there is a probability distribution over the time that the system stays in the same state and in case of change to which state the system changes.

## References

1. Aziz A, Sanwal K, Singhal V, Brayton R (2000) Model-checking continuous-time markov chains. *ACM Trans Comput Logic* 1(1):162–170
2. Baier C, Katoen JP (2008) Principles of model checking. MIT Press, Cambridge
3. Barnat J, Brim L, Krejci A, Safranek D, Vejnar M, Vejpustek T (2012) On parameter synthesis by parallel model checking. *IEEE/ACM Trans Comput Biol Bioinf* 9(3):693–705
4. Batt G, Page M, Cantone I, Goessler G, Monteiro P, de Jong H (2010) Efficient parameter search for qualitative models of regulatory networks using symbolic model checking. *Bioinformatics* 26(18):i603–i610
5. Benque D, Bourton S, Cockerton C, Cook B, Fisher J, Ishtiaq S, Piterman N, Taylor A, Vardi M (2012) BMA: visual tool for modeling and analyzing biological networks. In: 24th inter-

- national conference on computer aided verification. Lecture notes in computer science, vol. 7358. Springer, Berlin, pp 686–692
6. Bernot G, Comet JP, Richard A, Guespin J (2004) Application of formal methods to biological regulatory networks: extending thomas' asynchronous logical approach with temporal logic. *J Theor Biol* 229(3):339–347
  7. Biere A, Cimatti A, Clarke E, Fujita M, Zhu Y (1999) Symbolic model checking using SAT procedures instead of BDDs. In: Proceedings of 36th design automation conference, pp 317–320. IEEE Computer Society
  8. Bryant R (1986) Graph-based algorithms for Boolean-function manipulation. *IEEE Trans Comput C-35*(8):677–691
  9. Burch J, Clarke E, McMillan K, Dill D, Hwang L (1990) Symbolic model checking:  $10^{20}$  states and beyond. In: Proceedings of 5th IEEE symposium on logic in computer, science, pp 428–439
  10. Calzone L, Fages F, Soliman S (2006) BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics* 22(14):1805–1807
  11. Chabrier N, Fages F (2003) Symbolic model checking of biochemical networks. In: Computational methods in systems biology. Lecture notes in computer science, vol 2602. Springer, Berlin, pp 149–162
  12. Chabrier-Rivier N, Chiaverini M, Danos V, Fages F, Schächter V (2004) Modeling and querying biomolecular interaction networks. *Theor Comput Sci* 325(1):25–44
  13. Ciochetta F, Hillston J (2009) Bio-PEPA: a framework for the modelling and analysis of biological systems. *Theor Comput Sci* 410(33–34):3065–3084
  14. Clarke E, Emerson E (1981) Design and synthesis of synchronization skeletons using branching time temporal logic. In: Proceedings of workshop on logic of programs. Lecture notes in computer science, vol 131. Springer, Berlin, pp 52–71
  15. Clarke E, Grumberg O, Peled D (1999) Model checking. MIT Press, Cambridge
  16. Cook B, Fisher J, Krepska E, Piterman N (2011) Proving stabilization of biological systems. In: Verification, model checking, and abstract interpretation. Lecture notes in computer science, vol 6538. Springer, Berlin, pp 134–149
  17. Eén N, Sörensson N (2004) An extensible sat-solver. In: 6th international conference on theory and applications of satisfiability testing. Lecture notes in computer science, vol 2919. Springer, Berlin, pp 502–518
  18. Eker S, Knapp M, Laderoute K, Lincoln P, Meseguer J, Sönmez M (2002) Pathway logic: symbolic analysis of biological signaling. In: Pacific symposium on biocomputing, pp 400–412
  19. Fisher J, Henzinger T (2007) Executable cell biology. *Nat Biotechnol* 25(11):1239–1249
  20. Fisher J, Piterman N, Hajnal A, Henzinger T (2007) Predictive modeling of signaling crosstalk during *c. elegans* vulval development. *PLoS Comput Biol* 3(5):e92
  21. Guerriero M (2009) Qualitative and quantitative analysis of a Bio-PEPA model of the gp130/JAK/STAT signalling pathway. *Trans Comput Syst Biol XI* 5750:90–115
  22. Heath J, Kwiatkowska M, Norman G, Parker D, Tymchyshyn O (2008) Probabilistic model checking of complex biological pathways. *Theor Comput Sci* 391(3):239–257
  23. Kroening D, Strichman O (2008) Decision procedures: an algorithmic point of view. Springer, Berlin
  24. Kwiatkowska M, Heath J (2009) Biological pathways as communicating computer systems. *J Cell Sci* 122:2793–2800
  25. Kwiatkowska M, Norman G, Parker D (2007) Stochastic model checking. In: 7th international school on formal methods for the design of computer, communication, and software systems. Lecture notes in computer science, vol 4486. Springer, pp 220–270
  26. Kwiatkowska M, Norman G, Parker D (2008) Using probabilistic model checking in systems biology. *SIGMETRICS Perform Eval Rev* 35(4):14–21
  27. Moskewicz M, Madigan C, Zhao Y, Zhang L, Malik S (2001) Chaff: engineering an efficient sat solver. In: Proceedings of the 38th design automation conference, pp 530–535. ACM

28. de Moura L, Bjørner N (2008) Z3: an efficient smt solver. In: 14th international conference tools and algorithms for the construction and analysis of systems. Lecture notes in computer science, vol 4963. Springer, Berlin, pp 337–340
29. Pnueli A (1977) The temporal logic of programs. In: Proceedings of 18th IEEE symposium on foundations of computer science. IEEE Press, Piscataway, pp 46–57
30. Priami C, Regev A, Shapiro E, Silverman W (2001) Application of a stochastic name-passing calculus to representation and simulation of molecular processes. *Inf Process Lett* 80(1):25–31
31. Robinson A, Voronkov A (eds) (2001) Handbook of automated reasoning. Elsevier, Amsterdam
32. Schaub M, Henzinger T, Fisher J (2007) Qualitative networks: a symbolic approach to analyze biological signaling networks. *BMC Syst Biol* 1(1):4
33. Talcott C (2008) Pathway logic. In: Formal methods for computational systems biology. Lecture notes in computer science, vol 5016. Springer, Berlin, pp 21–53
34. Thomas R, Thieffry D, Kaufman M (1999) Dynamical behaviour of biological regulatory networks—I. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol* 55(2):247–276

# Chapter 11

## Computational Design of Informative Experiments in Systems Biology

Alberto Giovanni Busetto, Mikael Sunnåker and Joachim M. Buhmann

**Abstract** Accurate predictions of the behavior of biological systems can be achieved through multiple iterations of modeling and experimentation. In this chapter, we present the central ideas for the design of informative experiments in systems biology. We start by formalizing the task, and proceed by introducing the required tools to process data subject to uncertainty. We analyze design approaches which are Bayesian and information-theoretic in nature. A particular emphasis is placed on implicit and explicit assumptions of the available techniques. Two main design goals are here compared: reducing uncertainty and challenging existing belief. Finally, we discuss the limitations of the presented approaches to provide general guidelines for predictive modeling.

**Keywords** Hypothesis · DNA-damage · Causality · Deterministic · Activator-inhibitor · Goldbeter-Koshland function · Macromolecule · Phosphorylation · Aleatory variability

---

A. G. Busetto (✉) · J. M. Buhmann  
Department of Computer Science, ETH Zurich, Universitaetstrasse 6,  
8092 Zurich, Switzerland  
e-mail: busettoa@inf.ethz.ch

M. Sunnåker  
Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26,  
4058 Basel, Switzerland

A. G. Busetto · M. Sunnåker · J. M. Buhmann  
Competence Center for Systems Physiology and Metabolic Diseases, Schafmattstrasse 18,  
8093 Zurich, Switzerland

M. Sunnåker  
Swiss Institute of Bioinformatics, Universitaetstrasse 6, 8092 Zurich, Switzerland

## 11.1 Introduction

Biological systems are understandable at different scales and levels of detail [34]. Given appropriate data, mechanisms of interests can be modeled to perform accurate predictions [25]. A central question for experimentalists and modelers is:

*Which experiment should be selected to best answer a scientific question?*

For prediction, data quality matters more than quantity. Experimental design aims at selecting informative protocols for controlled experiments. Computational design entertains the idea that computers can help to maximize the task-relevant information gathered by the modeler through the measured data. The overall process of design typically involves multiple aspects, including those imposed by policy constraints and resource availability.

This chapter focuses on the task of experimental design from the theoretical and the computational point of view. To render the chapter self-contained, we start by providing a minimal set of preliminary notions. Expert readers may skip to the subsequent sections, which provide an overview of the fundamental principles. We proceed by describing the main set of goals and their interpretation. Assumptions, relations, and limitations of the approaches are discussed in the final section. Here, we focus on what is particularly relevant for systems biological applications. For concreteness, we introduce examples from the domain of cell signaling and biochemical network dynamics. This chapter is not an exhaustive dictionary of design techniques, but rather a comprehensive walk-through relating assumptions, goals, and limitations.

## 11.2 Preliminary Notions

Basic requirements for experimental design are:

- *hypothesis class*  $\mathbb{M}$ : the set of testable hypotheses;
- *experiment set*  $\mathbb{S}$ : the set of feasible experiments;
- *inference method*: the formal procedure employed to derive conclusions from the experimental observations.

Given a scientific question, hypotheses are formally expressed in terms of equations. In systems biology, equations are combined into systems that capture the essential behavior of components and interactions. Since the studied phenomena are typically time-varying, their models incorporate dynamic aspects to better predict the observed behaviors. Firstly, it is important to specify the model scope, that is the domain in which the model can be appropriately employed to perform predictions. Limitations might be experimental: for instance due to scarcity of resources, or because of the intrinsic inability to directly inspect the inner workings of the studied system. The experiment set conditions the choice of the hypothesis class, and vice versa. In principle, hypotheses and expected evidence should match the central prediction task,

which must be given a priori. Given the experimental data, conclusions are derived on the basis of the accumulated evidence. Some hypotheses are retained, others are discarded. Importantly, conclusions must include an estimate of the uncertainty associated with the selection of the hypotheses.

*Example Modeling DNA-Damage Mechanisms due to Irradiation.*

Let us consider the following process: the DNA-damage response pathway in mammalian cells. For illustration, let us take a model of p53/Mdm2 oscillations in response to ionizing radiation. The hypothesis class consists of two systems of equations. The two mechanisms model alternative hypotheses regarding the behavior of an oscillatory reaction network [17, 41]. Kinetic rate constants and other parameters are assumed to be known (with negligible uncertainty) from previous experiments. In this case, let us consider an experiment set containing two feasible experiments: high- and low-frequency gamma-irradiation of the cells. As damage is repaired, oscillations are counted as a function of irradiation time. Which experiment should be selected to maximally discriminate between the two putative mechanisms? Despite the complexity of the studied mechanisms, the design process is straightforward: there exist only two hypotheses (with known parameters) and only two experiments. In practice, real scenarios will involve large numbers of hypotheses, many possible experiments, as well as significant uncertainty with respect to rate constants and other parameters [14].

### ***11.2.1 Modeling of Dynamical Systems***

In its broadest interpretation, the formal process of modeling coincides with the hypothetico-deductive approach to science [37]. In practice, specific phenomena are modeled depending on the task. This is why it is not obvious to select the appropriate features of a phenomenon. The desired type of testable prediction induces different choices, which are ideally aligned with the aim of the modeling exercise.

Biological systems are time-varying processes. Hence, biological models typically involve time-varying entities. In systems biology, the emphasis is on variations due to mutual interactions between components. Predominantly due to (frequent) data scarcity, modeling might require the incorporation of previous knowledge from domain expertise and published results. Hypotheses are complex and require strong evidence for testing: armed with prior knowledge, the modeler may significantly improve predictions. Axiomatic assumptions and first principles are conveniently incorporated already in the definition of the model variables, for instance as state-space models [52].

Let us consider a state space  $\mathbb{X}$ . The state space describes all possible unique configurations of a process. The state  $x(t) \in \mathbb{X}$  is a variable which denotes the

instantaneous configuration of the studied process at time  $t$ . States could represent, for instance, the concentrations of several metabolites in a specific compartment of the cell, as well as more abstract entities such as stages of a cell cycle [54].

### 11.2.1.1 Process Model

Firstly, we take causality as essentially axiomatic.

**Main Assumption 1** *Causality.*

The future behavior of the studied phenomena can be described solely as the function of their current and past states.

This assumption excludes anticipatory effects and is in contrast, for instance, to batch image processing. Difference and differential equations constitute the classical formalism for modeling the behavior of dynamical causal processes. In the case of biochemical network analysis, such models well predict a large set of complex chemical interactions, including synthesis, binding, dissociation, degradation, allosteric activation, inhibition, and phosphorylation [17, 52]. More generally, these models can be employed to characterize the time-evolution of symbolic representations. All interactions between state components are captured by the system of equations. By doing so, the process dynamics is implicitly defined by the governing equations. The state variable is updated over time, generating trajectories which start with the initial conditions.

A classical example of state-space dynamical system is given by Ordinary Differential Equations (ODEs). This model class is widely used in systems biology [25]. In particular, the application of ODEs rests founded upon the established theory for deterministic modeling of biochemical reactions. Given the initial conditions, the *dynamical system*  $\mathcal{S}$  evolves as

$$\frac{dx(t)}{dt} = f(x(t), u(t), t, \theta), \quad (11.1)$$

where  $f$  is a function of

- $x(t)$ : current state;
- $u(t)$ : time-varying external input to the system;
- $t$  : time;
- $\theta$  : parameter vector out of parameter space  $\Theta$ .

In this particular example, the system is deterministic, memoryless (it satisfies the Markov property), and it evolves in a continuous state space. More generally, models may exhibit delays, memory, infinite-dimensional state spaces, stochastic behavior, and discreteness (in time or state space). Differential equations have been applied with success to a variety of applications. Yet they are not the only available modeling tool. Aside from practical limitations, there exist cases in which abstract and phenomenological representations might be appropriate as well. Regardless of whether



precise mechanisms are required or not, experimental design invariably aims at predictive modeling of the observable quantities. In practice, modeling might translate into finding estimates for  $f$  or  $\theta$ ; in some other cases, the modeler might be interested in predicting only the future evolution of the state given the data. In system identification, hypotheses  $M \in \mathbb{M}$  consist of specified functional forms  $f$  as well as their parameters  $\theta$ , that is  $M = (f, \theta)$ .

*Example Activator-Inhibitor.*

Let the state variable consist of two elements:  $x(t) = [R(t), X(t)] \in \mathbb{R}^{2+}$  (concentrations are non-negative quantities by definition). The elements  $R(t)$  and  $X(t)$  denote the concentration (for instance, in nM) of two time-varying macromolecules in a well-stirred and spatially homogeneous system at thermodynamic equilibrium [17]. There is a protein  $E$  in the system, which exists also in its phosphorylated form  $E_p$ . The macromolecule  $R$  is produced with a signal strength  $S$ , and  $R$  stimulates its own production by phosphorylating  $E$ .  $E_p$  stimulates the production of  $X$ , which in turn promotes the degradation of  $R$ . Precisely, the system equations takes the form:

$$\frac{d}{dt} \begin{bmatrix} R(t) \\ X(t) \end{bmatrix} = \begin{bmatrix} k_0 E_p(t) + k_1 S - k_2 X(t) R(t) \\ k_3 E_p(t) - k_4 X(t) \end{bmatrix} \quad (11.2)$$

$$f \left( \begin{bmatrix} R(t) \\ X(t) \end{bmatrix}, E_T, \cdot, \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} \right)$$

where the dynamics of  $E_p$  follows a Goldbeter-Koshland function [17]. Note the absence of time-dependence in this case: the system exhibits oscillations due to the interplay between  $R$  and  $X$  through  $E$  and  $E_p$ . If all parameters but one are known (e.g.,  $k_4$ ) the experimental design procedure could aim at optimizing the time points schedule. Ideally, this ensures that the time points are informative enough to obtain a good parameter estimate (for instance, by sampling above a certain frequency).

### 11.2.1.2 Measurement Model

Experiments are performed by the observer, often assuming negligible interference of the measurement apparatus to the behavior of the process.

**Main Assumption 2** *Non-interference.*

The measurement apparatus has a negligible effect on the behavior of the studied phenomenon (it does not change the state of the system).

This assumption is in contrast to cases in which quantum-mechanical effects cannot be neglected. Experimental observations can then be described in the data space  $\mathbb{D}$ , independently from the rest. The time-dependent readout variable  $y(t) \in \mathbb{Y}$  denotes the instantaneous measurement at time  $t$ . In terms of equations, the *measurement model* can be specified as follows:

$$y_\varepsilon(t) = h_\varepsilon(x(t), u(t), t, \eta, \xi), \quad (11.3)$$

where  $h_\varepsilon$  is a (generally nonlinear) function of state, input, time,

- $\eta$  are the tunable variables of the experimental protocol;
- $\xi(t)$  denotes the measurement noise, whose distribution is  $\mathcal{E}_t$ .

The variable  $\varepsilon \in \mathbb{S}$  denotes the choice of the experiment from the experiment set. *Time series data* are then acquired in the batch dataset

$$D_\varepsilon = \{(t_i, y_\varepsilon(t_i))\}_{i=1}^n \quad (11.4)$$

where  $n$  is the sample size. In principle, the noise distribution might be arbitrary: it could depend upon the state of the process, as well as on the input (for instance, in the case of stochastic interventions).

#### Example Measuring the Activator-Inhibitor.

Let us consider the activator-inhibitor process described in Eq. (11.2). The measurement process depends on the selection of the experiment, which is here denoted by the variable  $\varepsilon \in \mathbb{S}$ . For a fixed  $\varepsilon$ , assume that the concentrations of macromolecules  $R$  and  $X$  are not directly measurable, but the sum of their concentrations is. Thus,  $y(t) \in \mathbb{R} = \mathbb{Y}$ . The time-discrete measurement process introduces additive white Gaussian noise: for every time point  $t_i$ , the noise terms  $\xi_{t_i}$  are identically distributed and statistically independent from each other. The experiment variable could, for instance, control the parameters  $\eta$  (that is, mean  $\mu$  and variance  $\sigma^2$ ) of the normal noise distribution  $\mathcal{E}_t = \mathcal{N}(\xi|\mu, \sigma^2)$ . The measurement function  $h$ , which is time- and input-invariant, is linear with respect to both states and noise:

$$y_\varepsilon(t_i) = R(t) + X(t) + \xi_{t_i} = h_\varepsilon \left( \begin{bmatrix} R(t) \\ X(t) \end{bmatrix}, \cdot, \cdot, \begin{bmatrix} \mu, \sigma^2 \end{bmatrix}^T, \xi_{t_i} \right). \quad (11.5)$$

Assume, as in the previous example, that the goal is the estimation of the unknown kinetic rate  $k_5$  [see Eq. (11.2)]. In this scenario, the experiments indexed by  $\varepsilon$  with known  $\mu$  and small variance better reduce the uncertainty.

### 11.2.2 Modeling Uncertainty

Uncertainty arises through measurement noise, data scarcity, as well as from the impossibility of direct inspection of the inner workings of a system. It is possible to distinguish at least two types of uncertainty.

- *Aleatory variability*:  
due to the irreducible non-deterministic behavior of the process.
- *Epistemic indeterminacy*:  
due to the incomplete knowledge of the observer about the process.

The former depends only on the process, while the latter depends solely on the observer and on the available measurement apparatus. We assume that lack of complete knowledge has no effect on the behavior of the process, apart from the indirect effects induced by the selection of future interventions (such as inputs).

**Main Assumption 3** *Epistemic Separability.*

Excluding interventions, epistemic uncertainty does not effect the behavior of the observed process.

Yet the future behavior may be influenced by interventions selected on the basis of the current belief state of the observer. Degrees of plausibility for alternative hypotheses can be quantified in terms of belief states. The modeler may take advantage of this fact to design experiments in which, for instance,  $u(t)$  is a function of the residual uncertainty.

Belief states represent uncertain yet justified states of instantaneous knowledge. The belief states of an epistemic agent (that is, the modeler) are time-varying and depend upon the availability of new observations. As soon as new data are available, belief states can be updated to incorporate the additional evidence. Cox's theorem demonstrates that probability theory generalizes "common logic" (specifically, Aristotelian-Boolean logic) under uncertainty [19, 26]. Under weak assumptions, it can be shown that probabilities are the unique representations available to the modeler [20]. Informally, probability theory satisfies the following three desiderata: degrees of plausibility should [20]

1. be representable by real numbers;
2. agree with "common sense" (that is with basic Aristotelian syllogisms);
3. be consistent (epistemic agents with the same information must agree).

In this chapter, we subscribe to Cox's axioms of probability and to their interpretation as belief states for an epistemic agent. This framework is widely accepted, yet not the only possible choice, and Cox's assumptions are not undisputed.

Probabilities can be seen as frequencies of random, repeatable events but also as quantified uncertainties. They represent "risks" (that is, uncertainty about the occurrence of events specified within a stochastic model), as well as justified beliefs about hypotheses (models, parameters, etc.). As well as considering belief states for discrete sets, we also wish to consider the continuous case (which is typically the

case for the parameter space  $\Theta$ ). This goal is achieved by extending our discussion to probability densities.

*Example Uncertainty in Transcription.*

Control of transcription is a fundamental regulation mechanism in biology. The modeler considers the hypothesis class consisting of two candidate models of genetic regulation. The models describe mutually exclusive biochemical mechanisms [52]. Let us consider the following feasible set of experiments: concentration readouts of RNA-polymerase and of its binding frequency (to a certain promoter region). Each model consists of a set of Stochastic Differential Equations (SDEs), whose kinetic parameters are poorly known. The stochastic nature of the process is reflected in the aleatory uncertainty associated with the dynamics of the macromolecular concentrations. The hypothesis class consists of the two models: each model consists of the given system of equations, for all possible values which may be assigned to its parameters. Uncertainty about which set of SDEs should be selected, and with which specific parameters, is essentially epistemic (both before and after data acquisition); it only depends on the belief state of the observer. This is a case of model selection: the design aims at minimizing, for instance, the epistemic uncertainty associated with the selection of the “correct” model (that is, the most predictive model given the data).

### 11.2.2.1 Bayesian Inference

*Bayes’ theorem* is the application of the product rule between conditional probabilities<sup>1</sup>:

$$p(\mathcal{M} = M | \mathcal{D} = D) = \frac{p(\mathcal{D} = D | \mathcal{M} = M) p(\mathcal{M} = M)}{p(\mathcal{D} = D)}, \quad (11.6)$$

where

- the **hypotheses** are represented by the random variable  $\mathcal{M}$  whose sample space is the hypothesis class  $\mathbb{M}$  and
- the **data** are represented by the instance of the random variable  $\mathcal{D}$  whose data space is  $\mathbb{D}$ .

Bayes’s theorem relates three fundamental quantities:

- **prior**  $p(M)$ :  
the belief state for hypothesis  $M$  before the observation of the data;

---

<sup>1</sup> For simplicity, we simplify the notation  $p(\mathcal{X} = x)$  to  $p(x)$  when possible. In these cases,  $\mathcal{X}$  denotes the random variable and  $x$  an element of the respective sample space  $\Omega$ .

- **posterior**  $p(M|D)$ :  
the belief state for  $M$  updated after the observation of the data  $D$ ;
- **likelihood**  $p(D|M)$ :  
the probability of measuring the data  $D$  generated from hypothesis  $M$ .

The remaining term  $p(D)$  is the *evidence*. The evidence is a normalizing constant which can be calculated from prior and likelihood as

$$p(D) = \sum_{M \in \mathbb{M}} p(D|M)p(M). \quad (11.7)$$

Performing the normalization is in many cases a computational bottleneck: it involves non-trivial sums (or high-dimensional integrals).

In the Bayesian setting, the process of updating belief states when experimental evidence arrives constitutes the inference method. The posterior coincides with the obtained conclusions: it is determined by the likelihood and by the prior. The former term describes the relation between hypotheses and evidence. The latter term contains all information available a priori. To avoid prejudice, no presumed evidence is incorporated. To avoid bias, no arbitrary selection of the data is allowed.

**Main Assumption 4 Objectivity.**

Conclusions are derived based on all available evidence (and without presumed evidence).

During design, objectivity is a central step of simulated inference: experiments are evaluated according to the predicted outcomes of simulated datasets. Objectivity is so important, that it can be taken as a principle more than an assumption.

*Example Bayesian Estimation of Synthesis Rate.*

A simple motif of simultaneous synthesis and degradation can be obtained by combining basic rate laws:

$$\frac{dX(t)}{dt} = \underbrace{k_1 M(t)}_{\text{synthesis}} - \underbrace{k_2 X(t)}_{\text{degradation}} \quad (11.8)$$

where  $M(t)$  is the concentration of mRNA encoding protein  $X$ , while  $X(t)$  denotes the protein concentration, with given initial condition  $X(t) = 0$ . The function form of the ODE (that is, the governing function of the model) is known, yet the parameters are partially unknown. The degradation rate  $k_2$  is known (in arbitrary units), while the prior distribution of the synthesis rate  $k_1$  is exponential:

$$p(k_1) = \begin{cases} \lambda e^{-\lambda k_1}, & k_1 \geq 0, \\ 0, & k_1 < 0. \end{cases} \quad (11.9)$$

for a given  $\lambda$  (for concreteness, say  $\lambda = 1$ ). As an input,  $M(t)$  is directly controllable by the experimentalist. Let us consider a design scenario aiming at the minimization of a measure of uncertainty (for instance, variance) of the posterior distribution for  $k_1$ . The posterior distribution is given by the (normalized) product between the prior (in this case exponential) and the likelihood of the data. The likelihood function is defined by the measurement model through the distribution of the noise. The process is fully deterministic. As a consequence, there does not exist aleatory uncertainty for the elements of the hypothesis class. All uncertainty is epistemic and concerns solely the kinetic constant rates.

### 11.2.3 Measuring Information

Probability and information theory are deeply related [18, 26]. The former offers the framework to quantify uncertainty for individual hypotheses. The latter is based on the former and considers overall properties of belief states. The concept of self-information is the fundamental link between these frameworks. Informally, the self-information of an event corresponds to the “degree of surprise” associated with the particular outcome [18]. Events are specific observations of a random variable  $X$ . Formally, the *self-information* of the event  $E \subseteq \Omega$  is mathematically defined as<sup>2</sup>

$$h(E) = -\log_2 \left( \sum_{x_e \in E} p(x_e) \right) \quad (11.10)$$

where  $X$  is a random variable whose sample space is  $\Omega$  (with  $x_e$  of non-zero probability). Qualitatively, it is possible to note the following:

- events with *high* probability exhibit rather low self-information;
- events with *low* probability exhibit very high self-information.

Since both epistemic and aleatory uncertainties are modeled by probabilistic belief states, self-information measures the overall “surprise” of the observer for a particular readout (that is, due to the stochastic behavior of the process, as well as to the belief state of the observer).

---

<sup>2</sup> The choice of the logarithm depends on the (arbitrary) unit measure of information. We take the logarithm in base 2, and thus measure information in *bits*.

*Example Self-information of a Bistable Gene Network.*

Bistability is a feature often associated with systems exhibiting autocatalysis or positive feedback [30]. Let us consider a bistable synthetic single-gene autocatalytic network in *Escherichia coli* [6]. The process consists of a simple regulator and of a transcriptional repressor. A minimalist model for the process could be given by a system with two states (arbitrarily denoted as  $A$  and  $B$ ). The two states correspond to the combined instantaneous concentration of repressor and RNA polymerase in a single cell. Transitions between the two states are memoryless and stochastic. The system has known initial condition  $A$ . Unknown perturbations (for instance, uncontrollable external environmental stimuli) push the system back and forth from  $A$  to  $B$ . For the sake of simplicity, we assume that measuring the state exhibits negligible errors (that is, the estimation process is dominated by aleatory uncertainty). On the basis of previous experiments, the experimentalist knows that, state  $A$  is going to be measured with probability 0.9 despite the unknown perturbations. The self-information of the hypothetical readout  $A$  is thus approximately 0.15 [bits]. Measuring  $A$  is thus not very surprising. Conversely, observing  $B$  would yield approximately 3.32 [bits]. Indeed, high surprise is experienced after supposedly improbable events. When the event “the readout is  $A$ ” is almost certain (that is, with probability close to 1), its self-information is almost zero. Observing  $B$ , in contrast, would yield very high self-information (a significantly improbable event has been observed). Different results are expected when the system exhibits equal chances of being in state  $A$  or  $B$ . In such case, the self-information of the events would have been the same: exactly 1 [bit], like tossing a fair coin.

**11.2.3.1 Entropy as Uncertainty**

When a sender transmits the value of a random variable to a receiver through a noiseless channel, the average amount of communicated information is the *entropy*

$$H[p] = \mathbb{E}_p\{h(x)\} = - \sum_{x \in \Omega} p(x) \log_2 p(x), \quad (11.11)$$

that is the expected self-information over all possible outcomes.<sup>3</sup> A similar definition exists in the case of continuous random variables, for which the *differential entropy* is  $H[p] = - \int_{\Omega} p(x) \log p(x) dx$ . This limit is obtained partitioning the sample space into bins of progressively smaller width and by omitting the diverging term. In the case of biological experimentation, the outcomes are the measurement readouts.

<sup>3</sup> Note that  $\lim_{p \rightarrow 0} p \log p = 0$ .

Informally, the noiseless coding theorem states that the lower bound on the number of bits needed to transmit the state of a random variable is (asymptotically) given by the entropy [48]. This result is based, among other things, upon the assumption of a stationary source. The fundamental assumptions of classical information theory are formalized by the Shannon-Khinchin axioms [33]. Their discussion goes beyond the scope of this chapter, but it is important to highlight their interpretation to understand the analogy between communication and biological experimentation. In design, the source is the “state of nature”. Messages (that is, data) are sent through the noisy channel, which is the measurement apparatus. The receiver reconstructs the message (that is, the model) from the data. In brief, informative experiments correspond to channels with high bandwidth [13, 14].

Entropy is hence a functional of the belief state (in general, of a distribution). Peaked distributions yield low entropy, that we interpret as states of low uncertainty: very few hypotheses are considered to be plausible. As a measure of uncertainty, entropy is very general: it is consistent with other measures such as variance. For instance, in the case of a Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$ , differential entropy is given by

$$H[p]_{N(\mu,\sigma)} = \log(2\pi e\sigma^2)/2. \quad (11.12)$$

Since entropy is a measure of uncertainty, its reduction constitutes an admissible design goal. Indeed, entropy measures the expected depth of the shortest decision tree to identify the “correct model” given the data in the noiseless case. For illustration, let us consider this noiseless scenario. Let the hypothesis class consist of four equally plausible models. The modeler has to determine the correct model among the four solely on the basis of “yes/no” questions answered by an oracle (the oracle has complete knowledge). Let us identify the four models by their indexes ranging from 1 to 4. The modeler could, for instance, ask the following questions: is the index of the correct model smaller than or equal to 2? In case of positive answer, the effective hypothesis class would be restricted to models 1 and 2. In the next experiment, the modeler could ask directly whether model 1 is the correct model. In both cases, the correct model would be determined by the answer of the oracle. Similarly, if the oracle says that the index is not smaller or equal to 2, it would be worth asking whether it is 3. With these 2 conditional questions ((1, 2) Vs (3, 4)) followed either by (1) Vs (2) or by (3) Vs (4), all plausible hypotheses are considered. Knowing that the correct model has index smaller than 3 makes some questions redundant (it would be wasteful to ask if the model index is 4). This redundancy does not only exist due to assuming noise free measurements. In fact, probabilistic redundancy would also apply in the case of noisy oracles (whose yes/no questions cannot be fully trusted). In general, one might consider non-uniform plausibility for the possible models. In this case, how to avoid redundant questions? Some questions might be more informative than others (that is, able to yield a higher number of expected bits, which correspond to fewer plausible hypotheses). In principle, optimal questions are the ones which sequentially split the hypothesis class into sets of equal size (weighted according to



their cumulative probabilities). Then, what is the minimum number of such questions to determine the model? The entropy of the distribution over the hypotheses.

*Example Entropy of a Bistable Gene Network.*

Considering the bistable gene circuit of the previous example, it is possible to quantify the uncertainty of the experimentalist with respect to the state readouts. In the initial case in which  $p(A) = 0.9$ , the entropy amounts to approximately 0.47 [bits]. The entropy is maximal when  $p(A) = p(B) = 0.5$ , since the readouts are maximally indeterminate (1 [bit]). In contrast, when  $p(A)$  is almost certain, the entropy tends to zero, since an overwhelmingly small  $p(B)$  suppresses the (otherwise high) self-information of the event. For concreteness, let us consider the case in which entropy measures the uncertainty about the state of a dynamical system. The experimentalist aims at reducing the stochastic behavior of the bistable system to study other properties (so far not captured by the model). How could this be done? For instance this goal is achieved by introducing a forcing input as an intervention (that is, by controlling the state through  $u(t)$ ). Such a necessary step implements controlled laboratory conditions on a cell population. A set of interventions  $\varepsilon \in \mathbb{S}$  is then selected to minimize the aleatory state entropy.

Relative entropy is another fundamental information-theoretic quantity [38, 39]. Informally, it measures how many bits of information are (asymptotically) wasted when a sender communicates through a noiseless channel with the “wrong coding”. In this context, the coding is optimal when it achieves the highest bit-rate with respect to the probability  $p$  over the symbols generated by the source. Instead of using the optimal coding for  $p$ , coding is optimized for  $q$ . Already its functional form

$$R[p||q] = \sum_x p(x) \log_2 \frac{p(x)}{q(x)} \quad (11.13)$$

makes apparent its strict relation to entropy. Intuitively, it measures the following: how many bits are lost when  $q$  approximates  $p$ ? Relative entropy is particularly useful to evaluate approximations of probability distributions. At the same time, it can be used to measure the information gain between prior and posterior for given data [3].

### 11.2.3.2 Prior Knowledge

As we have seen, Bayesian inference requires priors. As long as experiments are independent, previous posteriors are the priors for further updates.

**Main Assumption 5** *Experimental Independence.*

Measurement outcomes of separate experiments are conditionally independent from each other.

This condition simplifies the calculation of Bayes' theorem. But there are cases in which this assumption is not satisfied. For instance, it might be difficult to guarantee independence when the same batch of cells is exposed to iterations of treatment and measurement. If initial conditions can be controlled and memory effects discounted, datasets  $D_\varepsilon^{(1)}$  and  $D_\varepsilon^{(2)}$  are assumed to be independent. The scenario induces the following decomposition of Eq. (11.6):

$$p(M|D_\varepsilon^{(2)}, D_\varepsilon^{(1)}) = \frac{p(D_\varepsilon^{(2)}|M)}{p(D_\varepsilon^{(2)})} \underbrace{\frac{p(D_\varepsilon^{(1)}|M)}{p(D_\varepsilon^{(1)})}}_{\text{prior for } D_\varepsilon^{(2)}} \overbrace{p(M)}^{\text{prior for } D_\varepsilon^{(1)}}. \quad (11.14)$$

The recursive nature of (conditionally independent) Bayesian inference makes it particularly convenient to perform incremental updates as soon as new data become available. This iterative information gain is a key feature of inference when uncertainty cannot be neglected. Just deriving point estimates would be insufficient, since such a description implies the loss of all residual uncertainty. In systems biology, this scenario is the norm, not the exception [43]. Consequently, a form of uncertainty propagation is necessary to perform a logically consistent analysis.

*Example* Two Experiments to Model a Multifunctional Enzyme.

Consider the case in which an enzyme catalyzes two reactions. A structural biochemical network is available in terms of ODEs. All kinetic constant rates are assumed to be known, except those of the two reactions (respectively denoted by  $k_1$  and  $k_2$ ). Assuming compatible experimental configurations and identical initial conditions, it is possible to separately obtain  $k_1$  from a first dataset  $D_\varepsilon^{(1)}$  and  $k_2$  from  $D_\varepsilon^{(2)}$ . Estimation results could be improved by simultaneously inferring  $k_1$  and  $k_2$ . This design requires the calculation of  $p(k_1, k_2|D_\varepsilon^{(1)}, D_\varepsilon^{(2)})$  instead of  $p(k_1|D_\varepsilon^{(1)})$  and  $p(k_2|D_\varepsilon^{(2)})$ .

So far we discussed cases in which priors are directly incorporated. We turn now to another important question: what can be done when prior probabilities are not available? Assigning zero priors to arbitrary hypotheses would impose zero posteriors as well. Irrespectively of any subsequent observations of the data some hypotheses would be unjustly excluded. No evidence would be able to modify the belief: the relative entropy between a non-zero posterior and a zero prior is infinite (it diverges). *Cromwell's rule* states that this issue should be avoided by assigning 0 and 1 prior probabilities exclusively for statements that are logically true or false (such as mathematical propositions) [40].

Yet there exist cases in which probabilistic reasoning has to be performed only on the basis of limited information. Transferring the available data from similar experiments might be challenging or practically unattainable (due to differences in experimental conditions, strains, etc.). Even worse, there might not be such data available to the modeler. What should be done in these cases? This is a delicate issue which deserves attention.

When the hypothesis class consists of a finite set of models, one may consider an external principle to assign epistemic probabilities. A common example is the *principle of indifference* [32]. Informally, the principle of indifference states that, given that only insufficient reasons exist to distinguish hypotheses, hypotheses should be considered equally plausible. Because probabilities are normalized to one, one obtains the uniform distribution over a finite hypothesis class. But what if the hypothesis class is not finite? Similarly, *non-informative priors* aim at exercising as little influence as possible on the posterior distribution. In the case of continuous parameter values with unbounded domains, such prior distributions may not be correctly normalized: they are called improper. Aside from difficulties due to normalization, the modeler has to be careful with difficulties arising from transformations of probability densities subject to nonlinear changes of variables [7, 8].

What can be said when only incomplete information is available? The *principle of maximum entropy* proposes a solution to this question [27, 28]. Degrees of belief are assigned according to the following rule: select among all constraint solutions the one which exhibits the largest entropy. Candidate distributions must be consistent with the available testable information  $T$ . In equations, when applied to priors it corresponds to

$$\text{select } p^*(M) = \arg \max_{p \text{ consistent with } T} H[p]. \quad (11.15)$$

Apart from setting priors, the principle may also be invoked for model specification [27]. Despite its arguable limitations, the principle of maximum entropy is not only mathematically satisfactory, but also epistemologically convincing [26]. It is consistent with the objectivity requirements. Informally, it states that no additional information should be presumed (that is, select the least committed and yet consistent belief). Non-maximum-entropy priors implicitly presume information which is not available, and consequently they support unjustified belief states. Testable information consist of any formal description which is amenable to statistical verification, such as known mean, variance, etc. Several well-known densities can be derived from maximum entropy considerations [8]: uniform (given finite support), exponential (given mean and non-negative support), Gaussian (given mean and variance), Laplace (given mean and expected variance), and famously Gibbs (given expected energy).

*Example Maximum Entropy Prior for Auto-regulation.*

Prokaryotic auto-regulation is a mechanism which can be modeled by a stochastic discrete dynamical system [52]. In brief, dimers of a protein repress their own transcription by binding to a regulatory region through a mechanism of auto-regulation. For the sake of simplicity, let us assume that the functional form  $f$  of the dynamical system is known exactly. Moreover, incomplete prior information for the parameters is available. Prior distributions are known explicitly (as posteriors from previous experiments) for all parameters except for  $k_r$ : the rate constant of binding to the regulatory region. Nonetheless, some testable information is available also for  $k_1$ . First, the constant is known to be non-negative (by construction). Furthermore, its expected value is  $\gamma$  (for instance, as found in the literature). On the one hand, a uniform distribution on the non-negative domain would yield an improper prior (because of lack of normalization) and would discard the available knowledge of  $\gamma$ . On the other hand, an arbitrary distribution would presume additional information which is not available (making it unjustified). The principle of maximum entropy yields instead the exponential distribution, as in Eq. (11.9) with  $\gamma = 1/\lambda$  for  $k_r$ .

## 11.3 Design Principles

Three general desiderata may be imposed on the design process: the designer has to

1. incorporate incomplete knowledge available a priori;
2. estimate costs and constraints associated with experimental procedures;
3. specify design goals.

The first desideratum is met by the application of Bayesian inference (or by other inference techniques) [15]. The second one is provided by external sources (domain knowledge, financing, other constraints). The experimental aim has to be explicitly stated by the scientist.

### 11.3.1 Goals

In essence, there exist at least two main design goals:

- **reducing** uncertainty [12],
- **challenging** existing belief (or theories) [13, 14, 21, 37].

Their difference is not only semantic: the goals correspond to different score functions for optimization. If epistemic uncertainty is measured in information-theoretic terms,

the former corresponds to minimizing (in expectation) the entropy of the posterior. The latter corresponds to maximizing (in expectation) the *information gain*, that is the relative entropy between posterior and prior. The information gain implicitly assumes that the posterior should be taken as the absolute reference to evaluate the quality of the prior. Entropy and information gain are functionals of the posterior, which depends on the data  $D_\varepsilon$  generated through  $\varepsilon$ . Respectively, the formal definitions of the score functions are as follows.

**Entropy minimization:**

select optimal experiment  $\varepsilon^*$  which minimizes the expected entropy:

$$\varepsilon^* = \arg \min_{\varepsilon \in \mathbb{S}} \mathbb{E}_{D_\varepsilon} \left\{ \underbrace{H[p(M|D_\varepsilon)]}_{\text{entropy}} \right\}. \quad (11.16)$$

**Information gain maximization:**

select optimal experiment  $\varepsilon^*$  which maximizes the expected information gain:

$$\varepsilon^* = \arg \max_{\varepsilon \in \mathbb{S}} \mathbb{E}_{D_\varepsilon} \left\{ \underbrace{R[p(M|D_\varepsilon) || p(M)]}_{\text{information gain}} \right\}. \quad (11.17)$$

In both cases, the expectation is taken over the evidence  $p(D_\varepsilon)$ , which is obtained from priors and (simulated) likelihoods, as in Eq. (11.7).

The goals coincide when the prior is uniform (since it has no effect on the posteriors). In this case, in fact, the following identities hold

$$\begin{aligned} \arg \min \mathbb{E} \{R[p(M|D_\varepsilon) || p(M)]\} &= \arg \max \mathbb{E} \left\{ \sum_M p(M|D_\varepsilon) \log_2 \frac{p(M|D_\varepsilon)}{p(M)} \right\} \\ &= \arg \min \mathbb{E} \{-H[p(M|D_\varepsilon)] - \log_2 |\mathbb{M}|\} = \arg \max \mathbb{E} \{H[p(M|D_\varepsilon)]\}, \end{aligned} \quad (11.18)$$

where  $|\mathbb{M}|$  denotes the cardinality of the hypothesis class. Furthermore, it is important to point out that maximizing the expected information gain is equivalent to maximizing the mutual information between models and data. Mutual information is a fundamental measure of statistical dependency [18]:

$$I[X, Y] = R[p(X, Y) || p(X)p(Y)]. \quad (11.19)$$

Finally, one has the following equivalence

$$\arg \max \mathbb{E} \{R[p(M|D_\varepsilon) || p(M)]\} = \arg \max I[M, D_\varepsilon]. \quad (11.20)$$

### 11.3.1.1 Requirements

The design requirements are as follows.

- Definition of hypotheses:
  - hypothesis class  $\mathbb{M}$  (containing the process models);
  - priors over the functional forms  $f$ , parameters  $\theta$ , the initial conditions.
- Definition of experiments:
  - set of experiments  $\mathbb{S}$ : it may consist of a joint set of measurable time points  $\{t_i\}_{i=1}^S$ , measurable components of the state space (their selection corresponds to the tuning of  $h$ ), as well as input interventions  $u(t)$  [13];
  - likelihoods (that is, the measurement model);
  - cost of each experiment.
- Choice of the design goal.

*Example* Design Scenarios for Modeling Tryptophan Biosynthesis.

There exist multiple alternative mechanistic models of tryptophan biosynthesis in bacteria [46, 53]. Let us assume that the hypothesis class consists of two functional forms  $f_1$  and  $f_2$ . They define alternative nonlinear system of ODEs (which could be of different complexity). The prior for the first hypothesis is 0.7 (hence, for the second it is 0.3). We denote this distribution as the pair (0.7, 0.3). The state space for the process model is defined by dimensionless concentrations of mRNA, enzyme, and tryptophan. Both initial conditions and parameters are partially unknown. Priors over parameters and initial conditions are given by the posteriors of previous compatible experiments. Subscribing to the principle of maximum entropy, exponential distributions are assigned when the expected value is known (since kinetic rate constant and concentrations are positive by definition), and uniform priors when a plausible range of values is given. Feasible experiments in  $\mathbb{S}$  measure relative concentrations of mRNA and tryptophan at a fixed number of time points. The design question is: if the number of time points is limited to 5, should the experimentalist measure every second ( $\varepsilon_1$ ) or every minute ( $\varepsilon_2$ )? The goal is to challenge existing prior belief. Thus, the expected information gain is the score function to maximize. For instance,  $\varepsilon_1$  could be preferable if it tends to reverse the order of prior probabilities more frequently than  $\varepsilon_2$  does. Indeed, with a (0.7, 0.3) prior, the information gain for the posterior (0.2, 0.8) (for fixed  $D$ ) is approximately 0.77 [bits], which is much larger than that obtained for the posterior (0.8, 0.2): approximately 0.04 [bits]. The effect of the prior is apparent in this case; the entropy of (0.2, 0.8) and (0.8, 0.2) is exactly the same.

In practice, design goals are constrained by resource availability. Otherwise, one would just measure as much as possible. The optimization formulations should take

this resource constraint into account; but how to relate costs  $c(\varepsilon)$  to benefits? The Pareto-efficient solutions are given by the *production-possibility frontier*, that is the set of experiments which dominate the others either in term of cost saving or of information [23]. In the design context, production corresponds to information gain (or uncertainty reduction) and possibility to the experiment costs. In practice, one typically asks directly one of the following questions. Given a certain maximal cost  $c_{\max}$ , which experiment yields the best results? This question would impose the cost constraint on the selection of the feasible experiments on the minimization/maximization goals of Eqs. (11.16, 11.17). Alternatively, one could ask: given this expected goal (in bits), which experiment minimizes the costs?

### 11.3.2 Calculation

The two main computational bottlenecks are belief update and optimization [14]. The former refers to the calculation of the posterior on the basis of prior distribution and of the uncertainty propagated through the system dynamics. The latter consists of the computational process to find the optimal solution (maximization of information gain or uncertainty reduction).

How to calculate the posterior  $p(M|D)$  [9]? Firstly, one has to propagate the uncertainty from one sample to another. Then, one has to update the propagated belief to incorporate new information [11]. For general dynamics, uncertainty propagation consists of calculating the solution of the Kolmogorov forward equation of the system [35] in the time interval  $[t_i, t_{i+1})$ . In the case of SDEs with a given  $f$ , it corresponds to calculating the solution of the Fokker-Planck equation [45]

$$\frac{\partial p_t}{\partial t} = \underbrace{\nabla \cdot [f p_t]}_{\text{drift}} + \underbrace{\Delta \Psi p_t}_{\text{diffusion}}, \quad (11.21)$$

which describes the time evolution of the time-varying distribution  $p_t = p(x|D_\varepsilon)$ . In the equation,  $\nabla \cdot$  denotes the divergence,  $\Delta$  is the Laplace operator with the diffusion tensor  $\Psi$ , while  $[f p]$  denotes the point-wise multiplication of  $p_t$  and the vector field  $f = f(x(t), u(t), t, \theta)$ . Generalized state-space propagation incorporates initial epistemic uncertainty of both initial conditions and parameter values into Eq. (11.21) [22]. On the right-hand side of the equation, the two terms denote drift (governed by the deterministic components) and diffusion (expressing the stochastic contributions). For deterministic systems described in terms of ODEs, as in Eq. (11.1), the diffusion term is absent: the propagated uncertainty is strictly epistemic. When the hypothesis class contains multiple functions, uncertainty propagation must be performed for each individual dynamical system and then normalized [14]. This significantly contributes to the computational challenge of the problem: closed-form solutions for the time-integrated trajectories rarely exist, and their numerical approxi-

mations are resource-demanding. Analogous procedures hold for dynamical systems whose state space and time are discrete.

After propagation, the update of the belief state is performed according to Eq. (11.6). This step typically involves the implicit calculation of the evidence for normalization. A large variety of techniques are available to jointly perform propagation and update [8, 9, 26]. Established techniques include Markov Chain Monte Carlo (MCMC) [24] and Kalman filtering [31]. Due to the hardness of the task, a number of approximations are available to the designer. Simulated inference may be achieved through variational approximations [8] or approximate Bayesian computation [51].

The quality of the overall optimization depends on the calculation of the updated solutions. Each evaluation of the score function involves the generation of hypothetical datasets and possibly the simulation of the respective updates. Depending on the goal and on the experiment set, the optimization might be in itself tractable (in the case of convex or submodular score function [36]) or very difficult (in the general case). There exist cases in which the designer can take advantage of the regularities of the score function [13, 36].

## 11.4 Discussion

We have seen assumptions, steps, and goals of computational design in systems biology. The framework that we have described is best understood from the Bayesian point of view, by employing probabilistic descriptions of justified belief states. As for other computational processes, design methods can be evaluated according to their correctness, efficiency, and simplicity of implementation. Here we highlight theoretical and computational aspects which deserve particular attention. Following are a list potential challenges and respective solutions for effective design.

- *Theoretical issuesx:*

- The choice of the **inference mechanism** is essential. The two main options available to the modeler are the Bayesian [15] and the classical frameworks [1]. It is worth noting that no unique Bayesian or classical viewpoint exist. Moreover, there has been much controversy about merits and limitations of each framework [8].

*Bayesian* approaches render the incorporation of prior information straightforward [14, 21]. Solutions consist of probability distributions over all hypotheses. This property is particularly appropriate for cases in which substantial uncertainty is expected to persist, as in biology [5]. The main limitation of this framework is its computational complexity: satisfactory approximations of posterior distributions might be unattainable.

*Classical* approaches are well established and sometimes able to offer computational shortcuts [4]. However, the incorporation of domain knowledge might be non-trivial: priors are often implicit.



- Which **prior** should be taken?

*Objective* priors require minimal assumptions. However, many uninformative prior are improper (and hence pathological) [8]. In principle, the Solomonoff-Levin distribution offers a general solution: it specifies the universal prior which is defined over the set of computable functions [49, 50]. At present, its applicability remains the subject of active research [44]. In practice, the two main principles available to the modeler are: indifference and maximum entropy. As a consequence, Gibbs, Gaussian and uniform densities are frequent choices.

*Subjective* priors are necessary when incorporating data from previous experiments. It often happens that (for mathematical or computational convenience) posterior distributions are specified in terms of “simple” forms such as categorical distributions or as members of the exponential family [8]. When only selected statistics of the distributions are available to the reader, subjective priors may be reconstructed from the available testable information according to the principle of maximum entropy.

- There exist a variety of possible **design goals**. Information theoretic approaches exhibit multiple benefits: they are general, rest on a well-founded theory, and enable the quantification of uncertainty on an absolute scale. Recent developments in information-theoretic model validation exhibit the potential to design experiments aimed at maximizing reliable information [10, 16].

*Uncertainty reduction* minimizes the epistemic uncertainty associated with the system.

The goal of *challenging existing belief* entertains the idea that information gains should be measured with respect to previous belief states [3]. As shown before, these two main goals coincide when prior information is unavailable.

- *Computational issues*:

- Particularly in the case of nonlinear dynamics, **uncertainty propagation** constitutes the dominant computational bottleneck. In practice, each evaluation of the score function might require the solution of the propagation task. Propagation can be directly combined with **updates**, taking advantage of the recursive nature of Bayes’ rule. Numerical solutions must be obtained through simulation when analytic solutions are not available (this is certainly the typical case in systems biological applications) [42]. It is important to note that parameter estimation and model selection by themselves are extremely demanding computational tasks; their discussion goes beyond the scope of this chapter (nonetheless the modeler should be aware of the impact of reliable estimation in the design process).

Exact solutions are rarely available. Kalman filters are commonly used for *linear* systems of ODEs.

If the belief state can be assumed to be *unimodal*, local linearization and Unscented Kalman filters offer excellent approximations [29]. Powerful sparse techniques are available as well [47].

Strongly nonlinear models constitute the norm in systems biology. Nonlinearities might induce *multimodal* belief states. Sequential Monte Carlo methods

are excellent to propagate general distributions, yet substantially resource demanding [11, 24]. Even checking their convergence might be non-trivial [11].

- **Optimization** is typically performed by taking advantage of some regularities of the score function, or by following established heuristics [2].

Entropy and mutual information may be *submodular* when conditional independence is satisfied [13, 36]. These cases are not only computationally tractable, but also efficiently solvable in practical applications [13].

General *nonlinear optimization* is required when arbitrary interventions are applicable to the systems (for instance, as inputs  $u(t)$ ). Very few general guarantees are available in the nonlinear case [2].

Systems biology is generally associated with large-scale data collection. Nonetheless, data quality matters: experimental design is an enabling technology for predictive modeling. In this chapter, we adopted a Bayesian information-theoretic viewpoint on experimental design. Computational design has the potential to assist human intuition in a key point: suggesting biologically interesting questions. We anticipate that the field of computational systems biology will move toward progressive automation of hypothesis generation and testing. In this context, design will play a crucial role to close the loop between modeling and experimentation.

## 11.5 Lessons Learned

- The analysis of many biological systems requires careful experimental planning due to limited time and financial resources.
- Methods for experimental design, in combination with mathematical models, provide means to assess the usefulness of potential (practically feasible) experiments. Such methods can be employed to reduce the experimental costs, since uninformative experiments are systematically avoided.
- Two common goals of experimental design are: uncertainty reduction and information gain maximization. In general, these goals may not be compatible; they do, however, coincide when all hypotheses are equally plausible a priori.
- Uncertainty propagation constitutes a significant computational bottleneck for the design of informative experiments aimed at dynamic modeling. However, careful assumptions on the distribution of the noise and Monte Carlo approaches may enable the construction of appropriate design schemes with acceptable computational costs.

## 11.6 Conclusions

The amount of biological data produced by high-throughput methods in biology has substantially increased in recent years. However, for many biological systems (e.g., signaling pathways) obtaining high-quality data involves a resource-intensive

process. Experimental design, in combination with computational methods for system identification, constitutes an important tool to reduce costs. This chapter primarily focuses on Bayesian approaches to experimental design; such framework relies on solid theory and established applications. The Bayesian point of view may prove particularly useful for the type of problems considered in systems biology: it is generally applicable, it is compatible with information theory, and there exist effective numerical approximation schemes which are already available to the designer. The discussed techniques exhibit the potential to accelerate the discovery of key principles and mechanisms of biological systems.

**Acknowledgments** We thank Sotiris Dimopoulos, Jörg Stelling, Cheng Soon Ong, Simonetta Scola, Gabriel Krummenacher, Alain Hauser, Elias Zamora-Sillero, Kay H. Brodersen, Jean Dautizneau, Andreas Krause and Marcus Hutter for insightful discussions and helpful comments. This work was financed with grants from YeastX and LiverX through the Swiss SystemsX.ch initiative, evaluated by the Swiss National Science Foundation.

## References

1. Atkinson AC, Donev AN (1992) Optimum experimental design. Oxford Science Publications, UK
2. Avriel M (2003) Nonlinear programming: analysis and methods. Dover Publications Inc., US
3. Baldi PF, Itti L (2010) Of bits and wows: a Bayesian theory of surprise with applications to attention. *Neural Netw* 23:649–666
4. Balsa-Canto E, Alonso AA, Banga JR (2008) Computational procedures for optimal experimental design in biological systems. *IET Syst Biol* 2(4):163–172
5. Bandara S, Schlöder JP, Eils R, Bock HG, Meyer T (2009) Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput Biol* 1:e1000558
6. Becskei A, Serrano L (2000) Engineering stability in gene networks by autoregulation. *Nature* 405:590–593
7. Berger JO (1985) Statistical decision theory and Bayesian analysis. Springer, Heidelberg
8. Bishop CM (2006) Pattern recognition and machine learning. Springer, Heidelberg
9. Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Wiley, New Jersey
10. Buhmann JM (2010) Information theoretic model validation for clustering. In: Proceedings of the 2010 IEEE international symposium on information theory, pp 1398–1402
11. Busetto AG, Buhmann JM (2009) Stable Bayesian parameter estimation for biological dynamical systems. In: IEEE CS Press proceedings of 12th IEEE international conference on computational science and engineering, pp 148–157
12. Busetto AG, Buhmann JM (2009) Structure identification by optimized interventions. In: Proceedings of 12th international conference on artificial intelligence and statistics, pp 57–64 (*J Mach Learn Res*)
13. Busetto AG (2012) Information theoretic modeling of dynamical systems: estimation and experimental design. Doctoral thesis, ETH Zurich, Zurich
14. Busetto AG, Ong CS, Buhmann JM (2009) Optimized expected information gain for nonlinear dynamical systems. In: Proceedings of 26th ICML, ACM series, pp 97–104
15. Chaloner K, Verdinelli I (1995) Bayesian experimental design: a review. *Stat Sci* 10(3):273–304
16. Chehreghani MH, Busetto AG, Buhmann JM (2012) Information theoretic model validation for spectral clustering. In: Proceedings of 15th international conference on artificial intelligence and statistics, pp 495–503 (*J Mach Learn Res*)

17. Conrad ED, Tyson JJ (2010) Modeling molecular interaction networks with nonlinear ordinary differential equations. In: Szallasi Z, Stelling J, Periwál V (eds) System modeling in cellular biology: from concepts to nuts and bolts. MIT Press, Cambridge, pp 97–123
18. Cover TM, Thomas JA (2006) Elements of information theory. Wiley, New Jersey
19. Cox RT (1961) The algebra of probable inference. Johns Hopkins University Press, Baltimore
20. Cox RT (1946) Probability, frequency, and reasonable expectation. *Am J Phys* 14:1–13
21. Daunizeau J, Preuschoff K, Friston K, Stephan K (2011) Optimizing experimental design for comparing models of brain function. *PLoS Comput Biol* 11(7):e1002280
22. Doucet A, Tadić VB (2003) Parameter estimation in general state-space models using particle methods. *Ann Inst Stat Math* 55(2):409–422
23. Fudenberg D, Tirole J (1983) Game theory. MIT Press, Cambridge
24. Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain monte carlo in practice. Chapman & Hall/CRC, US
25. Haefner JW (2005) Modeling biological systems: principles and applications. Springer, Heidelberg
26. Jaynes ET (2003) Probability theory: the logic of science. Cambridge University Press, Cambridge
27. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev Ser II* 106(4):620–630
28. Jaynes ET (1957) Information theory and statistical mechanics II. *Phys Rev Ser II* 108(2):171–190
29. Julier SJ, Uhlmann JK (1997) A new extension of the Kalman filter to nonlinear systems. In: Proceedings of aero sense: the 11th international symposium on aerospace/defense sensing, simulation and controls
30. Kærn M, Weiss R (2010) Synthetic gene regulatory systems. In: Szallasi Z, Stelling J, Periwál V (eds) System modeling in cellular biology: from concepts to nuts and bolts, pp 269–295. MIT Press, Cambridge
31. Kalman RE (1960) A new approach to linear filtering and prediction problems. *J Basic Eng* 82(1):35–45
32. Keynes JM (1921) A treatise on probability. Macmillan and Co., London
33. Khinchin AI (1957) Mathematical foundations of information theory. Dover Publications Inc., NY
34. Kitano H (2002) Computational systems biology. *Nature* 420:206–210
35. Kolmogorov A (1931) On analytical methods in the theory of probability. *Math Ann* 104:448–451
36. Krause A, Guestrin C (2005) Near-optimal nonmyopic value of information in graphical models. In: Proceedings of the 21st conference on uncertainty in artificial intelligence
37. Kuhn TS (1996) The structure of scientific revolutions. University of Chicago Press, Chicago
38. Kullback S (1959) Information theory and statistics. Wiley, London
39. Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22(1):79–86
40. Lindley D (1991) Making decisions. Wiley, London
41. Ma L, Wagner J, Rice J, Hu W, Levine A, Stolovitzky G (2005) A plausible model for the digital response of p53 to DNA damage. In: Proceedings of the National Academy of Sciences, vol 102, pp 14266–14271
42. Nelles O (2001) Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Springer, Berlin
43. Periwál V (2010) Bayesian inference of biological systems: the logic of biology. In: Szallasi Z, Stelling J, Periwál V (eds) System modeling in cellular biology: from concepts to nuts and bolts, pp 53–71. MIT Press, Cambridge
44. Rathmanner S, Hutter M (2011) A philosophical treatise of universal induction. *Entropy*. 13(6):1076–1136
45. Risken H (1996) The Fokker-Planck equation: methods of solutions and applications. Springer, Berlin

46. Santillán M, Mackey MC (2001) Dynamic regulation of the tryptophan operon: a modeling study and comparison with experimental data. *Proc Natl Acad Sci* 98(4):1364–1369
47. Seeger MW (2008) Bayesian inference and optimal design for the sparse linear model. *J Mach Learn Res* 9:759–813
48. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
49. Solomonoff R (1964) A formal theory of inductive inference, part I. *Inf Control* 7(1):1–22
50. Solomonoff R (1964) A formal theory of inductive inference, part II. *Inf Control* 7(2):224–254
51. Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2012) Approximate Bayesian computation in computational biology. *PLoS Comput Biol* (In press)
52. Wilkinson DJ (2006) *Stochastic modeling for systems biology*. Chapman & Hall/CRC, Boca Raton
53. Xiu ZL, Zeng AP, Deckwer WD (1997) Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. *J Biotechnol* 58(2):125–140
54. Zhong Q, Busetto AG, Fadedea JP, Buhmann JM, Gerlich DW (2012) Unsupervised modeling of cell morphology dynamics for high-throughput time-lapse microscopy. *Nat Meth* 9:711–713

# Chapter 12

## Predicting Phenotype from Genotype Through Reconstruction and Integrative Modeling of Metabolic and Regulatory Networks

Sriram Chandrasekaran

**Abstract** A central challenge in systems biology is the identification of molecular interactions that regulate organismal phenotype, and to predict phenotypic changes that arise from these interacting networks. The reconstruction of gene networks provides a mechanistic basis for understanding the genotype-phenotype relationship, and enables the simulation of cellular behavior resulting from genetic and environmental perturbations. Currently, there is a critical need for new methods that rapidly transform high-throughput genomics, transcriptomics and metabolomics data into such predictive network models for metabolic engineering and synthetic biology. This chapter describes tools and technologies that address these key challenges, with a focus on the algorithms, PROM and ASTRIX, which perform complementary functions in mapping and modeling gene networks. The Analyzing Subsets of Transcriptional Regulators Influencing eXpression (ASTRIX) approach builds Transcriptional Regulatory Networks from gene expression data while the Probabilistic Regulation of Metabolism (PROM) algorithm integrates disparate gene networks (metabolic and regulatory networks) together in an automated fashion. Some basic principles of reconstructing and modeling these networks are discussed, followed by a detailed description of these algorithms. Understanding how the networks function together in a cell will pave the way for synthetic biology and has wide-ranging applications in biotechnology, drug discovery and diagnostics.

**Keywords** Metabolic network · Transcriptional regulatory network (TRN) · Constraint-based modeling · Probabilistic regulation of metabolism (PROM) · Network inference · Analyzing subsets of transcriptional regulators influencing expression (ASTRIX)

---

S. Chandrasekaran (✉)

Institute for Systems Biology, 401, Terry Avenue North, Seattle, WA, USA  
e-mail: sriram@life.illinois.edu

S. Chandrasekaran

Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Champaign, IL, USA

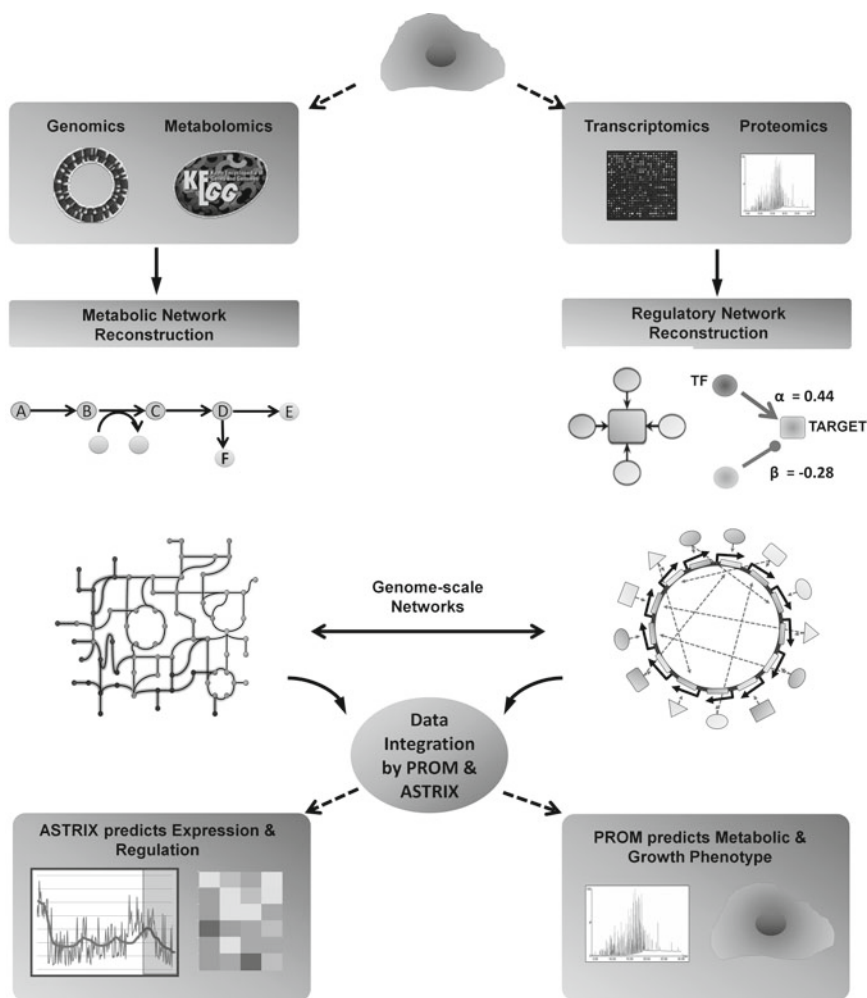
The genotype and the growth environment significantly influence the behavior and phenotype of an organism. Yet the mechanism of how a simple genetic change or environmental perturbation alters the behavior of an organism at the molecular level, and subsequently its phenotype, is still not completely clear. Systems biology aims to understand cellular behavior by identifying its molecular components and the interactions between them, and seeks to predict phenotypic changes that arise from these interacting networks. Systems biology primarily focuses on the entire system of interacting components ('networks') and predicts the emergent properties of these networks [22, 35, 41]. These 'cellular networks' are usually a group of genes or proteins that interact with each other and perform similar functions [4]. For example, the metabolic network carries out various bio-chemical reactions in each cell and the regulatory network controls these biochemical processes among others. Cellular behavior is determined by the differential activity of these networks; hence reconstructing and simulating these networks enables one to understand and better predict the response of a cell to an external perturbation. Figures 12.1 and 12.2 gives an overview of these different networks.

## 12.1 Reconstruction and Simulation of Metabolic Networks

Metabolism plays a central role in the functioning of an organism and is arguably the best understood cellular process. Yet, the size and complexity of the metabolic network poses a great challenge in modeling and simulating its behavior. Further, the lack of adequate data often limits our ability to test and analyze metabolism at the genome-scale using more traditional simulation methods such as reaction kinetics (modeled as a system of differential equations), where the mechanisms of reactions and their regulation is modeled individually and in detail.

Constraint-based modeling allows us to overcome such problems, because the only requirement is knowledge of the stoichiometry (the network topology) of the system in order to be able to fairly accurately simulate the potential metabolic behavior of an organism [45, 55]. By assuming steady-state kinetics, the system of differential equations required to model the system are simplified to a system of linear equations in constraint-based analysis. This technique thus requires much fewer parameters and can be applied to a wide variety of systems.

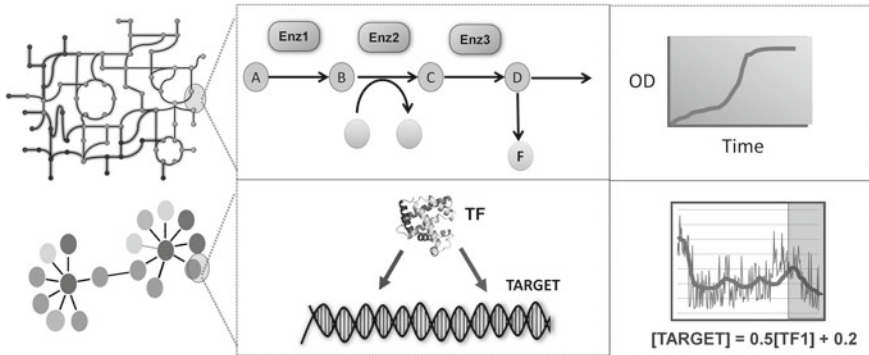
The biochemical network models built this way represent explicitly the mechanistic relationships between genes, proteins, and the chemical inter-conversion of metabolites within a biological system. Such genome-scale biochemical network models have been successfully completed for a variety of organisms including the prokaryote *E. coli* [53], eukaryotes such as *S. cerevisiae* [34, 51], and humans [21]. The reconstruction process has been greatly accelerated by the availability of online databases such as Kyoto Encyclopedia for Genes and Genomes [38] (KEGG) and SEED [32]. See Feist et al. [27] for a detailed description of the process of reconstructing, curating and validating these biochemical network models.



**Fig. 12.1** Graphical abstract: Generation and integration of high-throughput data to reverse-engineer cellular networks

In combination with physico-chemical constraints such as enzyme capacity, reaction stoichiometry, and thermodynamics, it is possible to determine the possible configurations in the metabolic network that correspond to physiologically meaningful states [54, 55]. Over the years, a number of methodologies have been developed to simulate the network, and these methods have enabled genome-scale analysis of microbial metabolism for various applications, from drug discovery to metabolic engineering, and modeling of microbial community behavior [28, 45]. The most prominent amongst them is flux balance analysis (FBA) [40]. FBA identifies the optimal flux pattern of a network that would allow the system to achieve





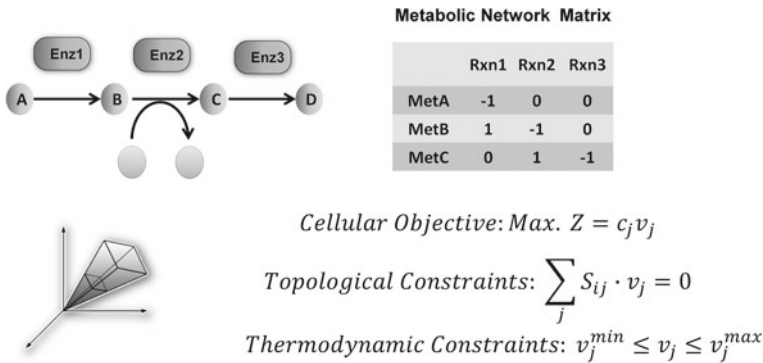
**Fig. 12.2** Metabolic and regulatory networks: Metabolism is at the heart of every cellular process, from energy production to producing precursors for processes like growth and cell division. Metabolic networks comprises of an array of enzymes that are involved in converting food into substrates for biosynthesis, or breakdown for energy production. Even a simple bacteria like *E. coli* has more than 2,000 biochemical reactions [53] that are involved in these processes. By simulating metabolic networks, one can predict an organism's phenotype such as growth rate and metabolic adjustments under diverse environmental conditions [6]. Transcriptional regulatory networks (TRNs, lower panel) are a specific kind of regulatory processes in a cell that are involved in controlling the expression of various genes in response to genetic and environmental changes. Modeling TRNs would hence enable the prediction of gene expression changes under various conditions [9, 14]

a particular objective, typically the maximization of an organism's growth rate or biomass production (Fig. 12.3).

Mathematically, FBA is framed as a linear programming problem:

$$\begin{aligned}
 & \text{maximize } Z = c_j v_j && \text{(the cellular objective)} \\
 & \text{subject to: } \sum_j S_{ij} \cdot v_j = 0 \quad \forall i && \text{(stoichiometric constraints)} \\
 & && v_j^L \leq v_j \leq v_j^U \quad \forall j \text{ (Thermodynamic constraints)}
 \end{aligned}$$

where  $i$  is the set of metabolites,  $j$  the set of reactions in the network,  $S_{ij}$  is the stoichiometric matrix,  $c_j$  is a vector that specifies which flux is being optimized (typically this is used for the maximization of growth) and  $v_j$  is the flux through reaction  $j$ . In genome-scale metabolic models of microbial systems, a biomass producing reaction is usually defined and used as the objective function. This reaction explicitly incorporates the chemical composition of the cell in terms of its macromolecules, nucleotide, amino acid, lipid and sugar content. These compounds are synthesized through an array of reactions that connect the input nutrients like glucose to the biomass components. Upper and lower bounds are placed on the individual fluxes based on thermodynamic considerations if they are available. For irreversible reactions, the lower bound  $v_j^L$  is set to be zero. Specific bounds, based on enzyme capacity measurements or thermodynamic considerations can be imposed on reactions; in the absence of any information these rates are generally left unconstrained



**Fig. 12.3** Constraint based analysis: A metabolic network is represented as a system of linear equations, represented in the form of a matrix (the Stoichiometric matrix). In the figure, A, B, C and D are metabolites involved in Reactions Rxn1, Rxn2 and Rxn3, catalyzed by enzymes Enz1, Enz2 and Enz3. This system is underdetermined, i.e., fewer equations than the number of variables. Therefore, we apply several constraints (depicted by a ‘flux cone’) to simulate various properties of the system. The reaction occurrence is limited by three primary constraints: reaction substrate and enzyme availability, mass and charge conservation, and thermodynamics. See text for more details on the constraints

i.e.,  $v^U = \infty$ , and  $v^L = -\infty$  for reversible reactions. To avoid unbounded solutions, i.e. Z reaching infinity, the input flux, typically the influx of glucose or other nutrients needs to be fixed to a specific value, and all fluxes should be viewed as relative to the input flux.

Changes to a metabolic network in vivo can arise not only due to perturbations to metabolic enzymes, but also due to transcription factors and other regulatory genes that control cellular metabolism. A major limitation of FBA is that it does not incorporate the effect of transcriptional regulation. Transcriptional regulation plays a central role in controlling metabolism and a key challenge in obtaining accurate predictions from biochemical networks is the integration of the gene regulatory network with the corresponding metabolic network [15, 17, 18]. The PROM algorithm addresses this issue and builds an integrated regulatory-metabolic network model. Before explaining how PROM works, we will first discuss how regulatory networks are reconstructed and modeled.

## 12.2 Reconstructing Transcriptional Regulatory Networks

Reverse engineering transcriptional regulatory networks, also known as regulatory network inference, involves the identification of functional modules or networks, which are a group of genes that regulate each other and perform similar functions [4]. Reconstructing the regulatory network enables one to understand the underlying molecular processes that cause phenotypic changes and better predict the response of

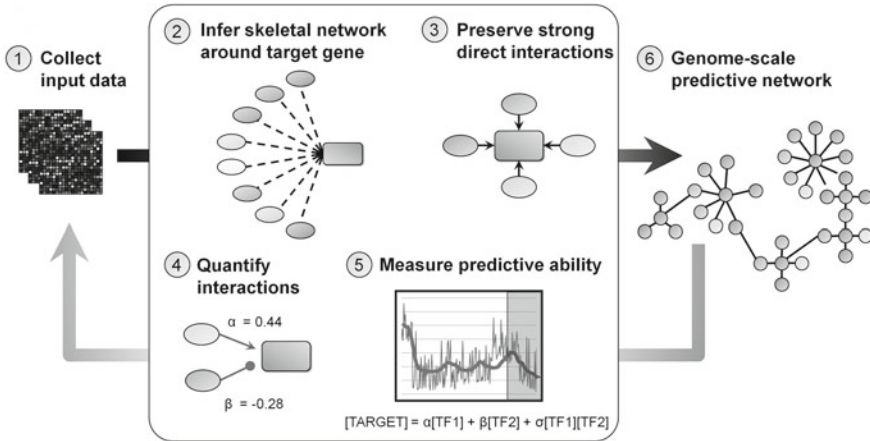
a cell to an external perturbation. Further, this would help us identify key hub genes that drive these networks and this knowledge plays a key role in drug discovery and synthetic biology.

There are hundreds of methods that build Transcriptional Regulatory Networks (TRNs), from binding [1, 31, 44, 60], gene expression data [9, 26, 47] or through integration of various data types [9, 49]. See [2, 4, 59] for a review of these various approaches. A commonly used approach has been to use omics technologies such as Microarrays and RNA sequencing, which provide a snapshot of the transcriptional activity of the cell [4]. Large repositories of gene expression data are currently available (GEO [23], M3D [25] and ArrayExpress [11]) enabling the rapid construction of genome-scale models of TRNs. Most of the expression-based ‘reverse-engineering’ methods primarily rely on the guilt-by-association principle—they try to identify functional relationships between genes by searching for similar expression patterns across diverse experimental conditions. The underlying assumption is that genes that have similar pattern of expression are generally co-regulated. For identifying transcriptional regulatory interactions, these patterns of co-expression are observed between all the transcription factors (TFs) and non-TFs in a cell. A gene is predicted to be regulated by a TF if they share a significant similarity in their expression patterns. This ‘similarity’ of expression is easy to understand but usually hard to measure. A suite of methods that use different metrics ranging from correlation and mutual-information to least-angle regression have been developed to infer similarity in expression, with each one having its own advantages and limitations. See [4, 8, 47, 48] for a review of these inference algorithms.

## 12.3 Inferring TRNs Using ASTRIX

Analyzing Subsets of Transcriptional Regulators Influencing eXpression (ASTRIX) combines two well-known inference algorithms that use disparate principles—ARACNE (Mutual information) and LARS (Regression)—to infer TRNs. Briefly the novelty of ASTRIX is that it not only infers interactions between TFs and target genes, but also creates a predictive model of the network which can be used to predict expression in new conditions. Also, unlike most approaches that infer networks for all the TFs and target genes, ASTRIX takes into account the limitation of the provided data set and infers a network structure only for a subset of genes that could be modeled well with the provided expression data.

ARACNE is a mutual-information based method for identifying transcriptional interactions between genes using gene expression data. Like correlation, mutual information is a metric that detects statistical dependence between two variables, but unlike correlation, it does not assume linearity, continuity, or other specific properties of the dependence. Information-theoretic approaches are comparatively effective for studying large networks where putative gene-gene interactions are learned from a relatively small amount of expression data [12]. ARACNE predicts a gene and transcription factor to interact if the mutual information between the expression



**Fig. 12.4** The ASTRIX approach for reverse engineering transcriptional regulatory networks (TRN): Each step (numbered) is explained in the text

levels of the gene and its potential regulator is above a set threshold. ARACNE has been shown to accurately reconstruct the regulatory network of c-Myc in B-cells [5] and has recently been used to reconstruct the TRN responsible for epithelial to mesenchymal transformation in Gliomas [13]. One of the main novelties of the ARACNE algorithm is that it uses the Data Processing Inequality (DPI) technique to eliminate the majority of indirect interactions inferred by co-expression methods.

Least Angle Regression (LARS) is a regression algorithm used for inferring relationships between a dependent variable (response, in our case gene expression) and one or more independent variables (predictors, TFs), and also for prediction and forecasting the state of the response variable. LARS is a model selection algorithm, similar to, but a less greedy version of the traditional forward selection method. It selects a parsimonious set of predictors from a large collection of possible covariates for the prediction of a response variable [24].

ASTRIX combines LARS and ARACNE, and uses genome-scale gene expression data to infer a transcriptional regulatory network model capable of making quantitative predictions about the expression levels of genes given the expression values of the transcription factors [14]. The ASTRIX algorithm works as follows: we first compile a large set of microarray data for the particular system of interest. Ideally, the data should measure the transcriptome of the system under various perturbations and environmental conditions (step 1 in Fig. 12.4). Generally, any reverse engineering method requires the availability of a large set of gene-expression data that profiles a broad range of cellular states and associated gene-expression levels [5, 25]. This is necessary because genetic interactions are best inferred when the genes explore a substantial dynamical range [5, 26]. In simple organisms, a wide range of states can be sampled by large-scale gene knockouts or environmental constraints. On the other

hand, these data might not be easily available for more complex cellular systems and the network inference algorithm has to utilize the naturally occurring phenotypic variations to reverse engineer the cellular network [5]. For each gene in the system, a “skeletal” network is inferred around the target gene using ARACNE [50] (step 2), and DPI is used to eliminate indirect interactions. A stringent mutual information threshold is chosen ( $p\text{-value} < 10^{-6}$ ) so that only strong interactions are retained (step 3). This serves to identify key regulators that share a high degree of mutual information with the target gene of interest. We then use the transcription factors or regulators predicted for each gene by ARACNE and fit a regression model using LARS [24] (steps 4 and 5). This would take the form:

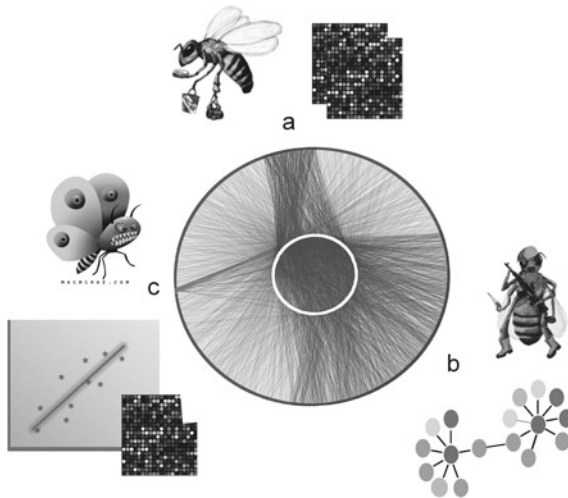
$$[\text{Target Gene}] = \alpha[\text{TF1}] + \beta[\text{TF2}] + \dots + \sigma$$

where  $\alpha$ ,  $\beta$ , and  $\sigma$  are constants inferred by LARS based on gene expression data. The accuracy of the model inferred by LARS can be determined by measuring the Root Mean Square Deviation (RMSD) of the model prediction with the actual expression. All data used in this procedure are normalized before network inference to have row variances of 1. Thus, for a given influence of a transcription factor on a given target gene, one can uniformly interpret the magnitude of the coefficients  $\alpha$ ,  $\beta$ , and  $\sigma$ , and use their magnitude to rank the individual interactions [9]. Also, since RMSD has the same units as variance, we get a clear interpretation of the amount of variance of the gene explained by the model. This process is repeated for all of the genes in the dataset or a subset of interest (step 6).

As mentioned earlier there are many advantages of using ASTRIX over using ARACNE or LARS alone. While ARACNE gives only the topology of the network, ASTRIX gives both the topology and also predicts if each interaction is activating or inhibitory. Most importantly, ASTRIX only selects the subset of TF—gene relationships that can accurately predict the target gene’s expression quantitatively in the training set, and only these interactions are moved forward to the test set for validation. Recent analyses have shown that by using a combination of different approaches, like mutual information and regression, is more effective at inferring networks than the individual methods used alone [47]. One could thus expand our framework by using other method types like Bayesian networks [29] and Random Forests [36] to infer interactions. Further, combination with other data types like binding, sequence and TF knockout data, if available, will lead to more comprehensive and predictive models of transcriptional regulation [49].

## 12.4 Application: Inferring the Honeybee Brain TRN for Social Behavior Using ASTRIX

The natural behavior repertoire of the honeybee (*Apis mellifera*) is perhaps the best studied of any non-human animal [67]. They exhibit complex social behaviors like aggression, nursing, foraging and spatio-temporal learning, which are influenced by both genetic and environmental factors [14, 57]. Further, these behaviors are



**Fig. 12.5** Overview of the approach used to build a TRN model for the honeybee brain. **a** Brain expression from different bee brain behavior states were collected. **b** A network model was built using ASTRIX, and **c** the network model was then used to predict expression in new conditions and identified key regulators of specific behavior processes. The network wheel in the center displays the inferred TRN model. The middle circle has the TFs and the outer circle has the 2,176 target genes. Darkly shaded edges are interactions between TFs and target genes involved in specific behaviors like aggression, foraging or maturation (indicated by the bees). ASTRIX identified TFs unique to each behavior and global regulators that controlled multiple behaviors

dynamic and associated with multiple levels of cognitive and molecular processing [58]. Given its dynamical nature, it's not known if behavior is influenced by the kind of Transcriptional Regulatory Networks (TRNs) known to regulate other phenotypes like development [20, 42, 52]. We hypothesized that behaviorally related brain gene expression could be used to reconstruct the type of transcriptional regulatory networks (TRNs) that operates for other phenotypes.

To enable comprehensive network inference, the bees were sampled in one of 48 different states defined by behavior, genotype, and environment. Nearly all genes expressed in the bee brain were differentially expressed in at least one of the 48 states; this broad survey captured natural variation across most of the transcriptome, even without experimental genetic perturbation. We then constructed a regulatory network for the honeybee brain that identifies key regulators that control the genes that are responsible for the phenotypic changes (Fig. 12.5).

The ASTRIX algorithm was then used to infer a predictive transcriptional regulatory network model for the honeybee brain using this data from 2,000 microarrays involving 48 different behavior states. ASTRIX accurately predicted the expression of 2176 genes involved in these behaviors with an average correlation of 0.87 in test conditions based on the expression of transcription factors. It identified transcription factors that are central actors in regulating behavior in the honeybee brain, and

our results suggest a remarkably close relationship between brain transcriptome and behavior. One can draw a picture where a core module of transcription factors is responsible for various phenotypic changes.

ASTRIX's ability to predict expression even in new phenotypes suggests a relatively complete and accurate reconstruction of the transcriptional regulatory network underlying these changes. Accurate prediction of quantitative behavior is the ultimate test of our understanding of a given system, and will enable re-engineering of cellular circuits. Our ability to model a surprisingly high percentage of the transcriptome, without information on physical interactions or brain subregion localization—implies that the relationship between brain gene expression and behavior is both stronger and more predictable than previously imagined. A more detailed analysis and description of the study can be found in [14].

## 12.5 Integration of Regulatory and Metabolic Networks Using PROM

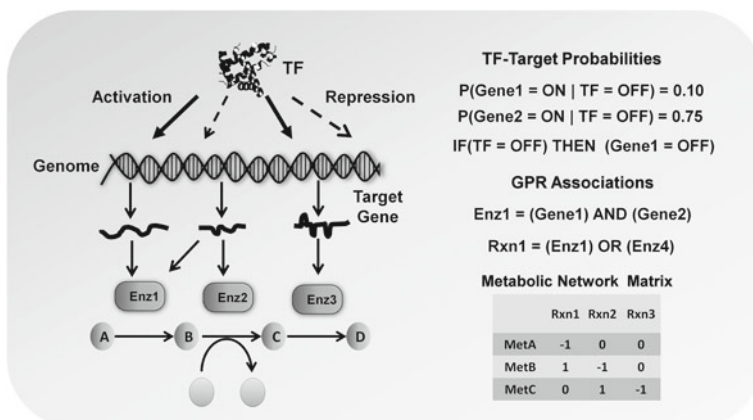
Till now we have discussed methods to infer and model TRNs. A forefront challenge in modeling organisms today is to build integrated models of regulation and metabolism. Predicting the effect of transcriptional perturbations on the metabolic network can lead to accurate predictions on how genetic mutations and perturbations are translated into flux responses at the metabolic level. Further, it can assist in the engineering of genetically modified organisms for synthetic biology and drug discovery [56]. Studying the molecular networks that distinguish a normal cell from diseased one may lead to the identification of critical metabolic biomarkers for cancer and other diseases [65]. Although gene and protein expression have been extensively profiled in human diseases, little is known about the global metabolomic changes that occur due to these perturbations. Profiling such metabolic alterations can lead to the discovery of metabolic markers [65].

## 12.6 Challenges of Integrated Regulatory-Metabolic Network Modeling

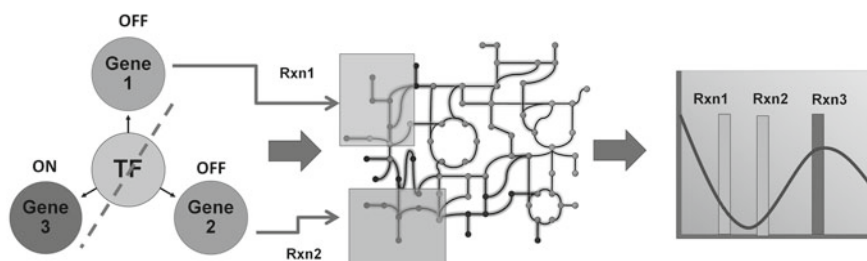
From our description of the ways to reconstruct and model the metabolic and regulatory networks, it is clear that the two network types have very different ways of being modeled. While the TRNs are simulated based on statistical associations, the metabolic networks are modeled based on a detailed biochemistry based mechanistic framework and constrained by thermodynamics, mass and energy balance (Fig. 12.6). So it has been a difficult challenge integrating different modeling paradigms.

Many methods solve this problem of network integration indirectly, by using gene expression data, which is the output of transcriptional regulation, with the metabolic





**Fig. 12.6** Overview of Integrated modeling: The figure highlights the way in which the metabolic and regulatory networks can be represented in silico and integrated together. (GPR—Gene-Protein-Reaction)



**Fig. 12.7** Overview of RFBA: The deletion of a TF results in alteration in expression of its target genes. These are then mapped onto the corresponding reactions in the metabolic network. If the target gene is determined to be OFF, the fluxes through the reactions are turned off and the optimal flux state (*curved line*) and the growth rate is determined using FBA. Note that in RFBA, genes and fluxes can only have two states (ON/OFF)

network [7, 16, 37, 63]. While these are very effective, they do not explicitly account for transcriptional regulation and cannot simulate perturbation to transcription factors. The first successful integration of these network models was by the algorithm RFBA [17, 18]. The RFBA approach and its variants [19, 64] simplify regulation to an ON-OFF process, instead of the complex quantitative models that are used to model TRNs. Figure 12.7 describes the RFBA model. Briefly, the states of genes in the metabolic model are determined by transcription factors using manually constructed regulatory rules. For example the state of a metabolic gene might be given by  $A = \text{TF1 AND TF2}$ , which means that both TF1 and TF2 should be in ON state for gene A to be active.

Although these Boolean logic based interactions are easy to understand, usually regulatory interactions are more complex than a binary process where genes and



reaction fluxes can only have two states in the population—ON or OFF. As these are built by manual curation of literature, given the large number of interactions, it is extremely difficult and time-consuming to manually write Boolean rules and identify significant interactions at the genome scale using RFBA. Due to these reasons there have been very few such integrated models available in literature [17, 30, 33].

## 12.7 Probabilistic Integrative Modeling Using PROM

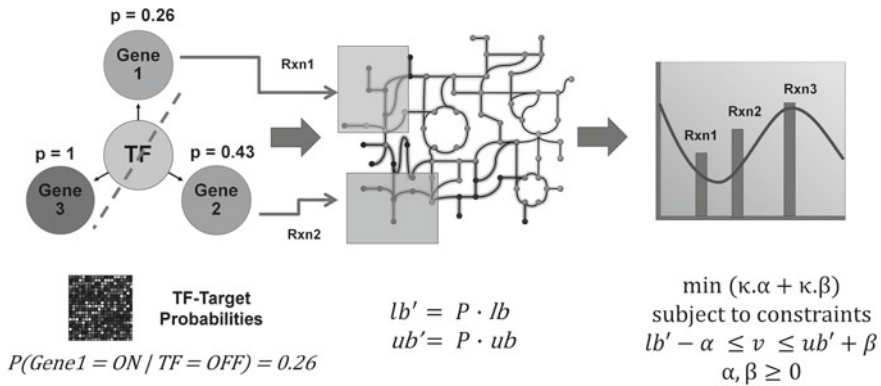
To overcome these drawbacks, we developed the PROM approach that addresses these issues. PROM, enables direct integration of the transcriptional and metabolic networks for modeling and overcomes the need for manually writing the Boolean rules by automatically quantifying the interactions from high throughput data—thereby greatly increasing the capacity to generate genome-scale integrated models. The model framework, based on constraint-based analysis, is designed to circumvent the need for kinetic parameters for metabolic modeling, and most importantly does not assume direct correlation between enzyme activity and mRNA expression.

PROM's novelty lies in the introduction of probabilities to represent gene states and gene-transcription factor interactions. PROM can algorithmically quantify these interactions based on microarray data. For example, the probability of gene A being ON when the regulating transcription factor B is OFF is given by  $P(A = 1|B = 0)$ ; similarly  $P(A = 1|B = 1)$  gives the probability of A being ON when B is ON. The relationship between TF and target gene is then quantified using microarray data. So, if we estimate that in 80 % of the samples we find the gene to be ON, and 20 % of the samples it is OFF or not expressed, then the probability associated with a gene being ON is 0.8. Once this interaction information has been defined, one can model the effect of perturbations to the regulatory network on the metabolic network using PROM (Fig. 12.8).

To model the effect of a TF knockout at the genome scale, the states of all its regulatory target genes are determined. These probabilities are then used to constrain the fluxes through the reactions controlled by the target genes.

$$lb' = P \times lb^*; ub' = P \times ub^*$$

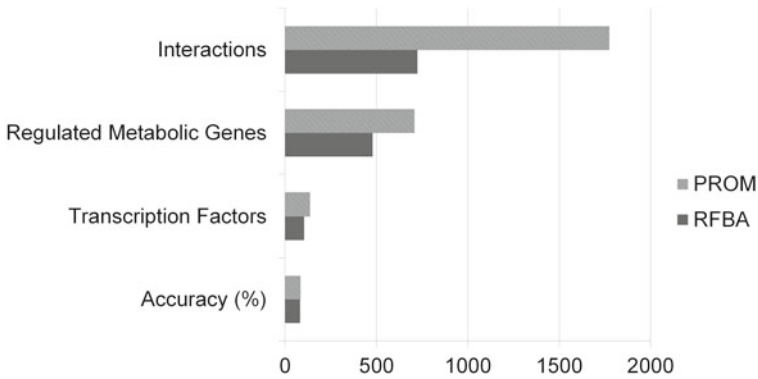
$ub^*$  and  $lb^*$  are the maximum and minimum fluxes through a reaction and these are determined for each reaction using Flux Variability Analysis [46] on the unregulated metabolic model, and  $P$  is the probability defined previously. To account for other factors that may affect enzyme activity such as translational, post-translational and metabolite interactions, these constraints are used as cues to determine the most likely flux through a particular enzyme. Unlike thermodynamic or environmental constraints that cannot be violated, the regulatory constraints are 'soft' constraints, so the system can exceed these constraints to maximize growth, but with a penalty. The magnitude of this penalty factor ( $\kappa$ ) determines the strength of the transcriptional



**Fig. 12.8** Overview of the process used to integrate the metabolic and regulatory network using PROM. The Transcription factor (TF) states are determined based on environmental conditions; the state of TF is then used to determine the ON/OFF state of the target genes based on probabilities estimated from microarray data. The probabilities are then used to constrain the fluxes through the metabolic network. The optimal metabolic state (*curved line*) that satisfies both the thermodynamic and regulatory constraints is determined. In PROM, the constraints based on gene expression are used as cues to obtain the optimal flux state. Note that unlike RFBA, where genes and fluxes can only have two discrete states (ON/OFF), PROM can have a more continuous restriction of flux levels. Further, PROM's automated inference of interactions and probabilistic formalism enables it to create comprehensive models.  $lb'$ ,  $ub'$  are constraints based on transcriptional regulation,  $\alpha$ ,  $\beta$  are positive constants which represent deviation from those constraints (determined by the solver) and  $\kappa$  is the penalty factor

constraints, with higher values implying stronger effect of regulation. Following this procedure, we arrive at an optimal model, which satisfies most or all of the regulatory constraints (Fig. 12.8).

The construction of an integrated metabolic-regulatory network using PROM requires the following: (1) the reconstructed genome scale metabolic network (2) regulatory network structure, consisting of transcription factors (TF) and their targets; (3) gene expression data. PROM then overlays these regulatory interactions on top of the metabolic network, which as mentioned earlier is now available for a large number of organisms. PROM utilizes the Gene-Protein-Reaction (GPR) relationships present in the metabolic network models to connect the regulatory targets (obtained from literature, or from databases like RegulonDB, Yeastract [61, 66] or inferred using network inference approaches like ASTRIX) to the corresponding metabolic reactions. The GPRs takes into account the presence of isozymes or multi-gene, multi-subunit complexes that may be involved in each metabolic reaction. We can then test the performance of the integrated model by simulating growth phenotypes under different environmental conditions. Figure 12.9 compares the integrated network models from PROM and RFBA for *E. coli*.



**Fig. 12.9** Comparison between PROM and RFBA. Using PROM, integrated regulatory-metabolic networks for the model organism *Escherichia coli* was constructed and we predicted the growth phenotypes of 15 TF knockouts in 125 different growth conditions with 85 % accuracy. Note that PROM based on automated inference is as accurate as the manually curated RFBA model

## 12.8 Application: Constructing an Integrated Network Model for Tuberculosis Using PROM

After validating the approach, PROM was used to build the first genome-scale integrated metabolic-regulatory model for *Mycobacterium tuberculosis*, a critically important human pathogen. PROM was specifically designed to be applied to less-studied systems like *M. tuberculosis*; by integrating various high throughput data, PROM can help us understand the system in a more holistic manner. The regulatory data for *M. tuberculosis* was compiled by Balazsi et al. [3] and gene expression data consists of 437 whole-genome microarrays of *M. tuberculosis* measuring the effects of 75 different drugs [10].

We systematically knocked out all the TFs in *M. tuberculosis* that regulate metabolic genes, and their knockout phenotypes were predicted using PROM. Comparison with gene knockout data [62] revealed that PROM predicted the phenotypes with an accuracy of 95 %. PROM also identified key genes that regulate vital steps in metabolism, which could lead to the prediction of better drug targets for therapy. Indeed, out of the 11 predicted essential genes by PROM, 7 of them were drug targets, which is highly significant ( $p$ -value = 0.01). Despite the lack of complete biological knowledge about *M. tuberculosis*, PROM was still able to predict the phenotypes with relatively high accuracy.

PROM represents the first automated integration of a genome scale TRN with a biochemically detailed metabolic network, bridging two important classes of systems biology models that are rarely combined quantitatively [15]. Several challenges will need to be addressed to build integrated regulatory-metabolic models for systems in higher organisms. While our models have shown great accuracy to date for simple organisms, we have not yet demonstrated their success in human systems, where the

complexity of regulation encompasses not only the effect of transcription factors, but also the effect of non-coding RNAs, epigenetic effects, post-translational modifications, and alternative splicing. With the development of methods that incorporate other network types, like signaling [17, 43] and a range of other cellular processes [39], one can envision transitioning these models to higher systems. Further, with the advent of automated approaches for metabolic network reconstruction [32], integrated network models could be constructed rapidly for a wide array of sequenced organisms.

## 12.9 Conclusion

Despite recent advances in computation, new algorithms are needed to integrate various data sources, and to assemble a holistic view of the cell. The new approaches discussed here address this issue and have diverse applications for understanding microbial biochemistry, drug discovery and disease progression. ASTRIX can identify key hub genes that drive networks, which could aid in synthetic biology, and also for finding drug targets against both microbes and cancer cells. Further, predicting the effect of transcriptional perturbation on the metabolic network using PROM can lead to more effective metabolic engineering of microbes and the identification of critical metabolic biomarkers for cancer and other diseases.

## 12.10 Lessons Learnt

Reconstructing and integrative modeling of metabolic and regulatory networks allows one to better understand the genotype to phenotype relationship, and paves the way for metabolic engineering and synthetic biology. Emerging tools and algorithms that integrate diverse high throughput data and build genome-scale models of these networks were discussed. The ASTRIX algorithm [14] allows the reverse-engineering of regulatory network models from high throughput data. ASTRIX identifies key hub genes that control cellular networks, and these network models can quantitatively predict gene expression changes in new conditions. The PROM algorithm is an ideal tool for constructing genome-scale regulatory-metabolic network models in an automated fashion [15]. Using PROM, the first integrated genome scale model for the pathogen, *M. tuberculosis*, was constructed. Furthermore, PROM can detect drug targets and metabolic flux changes, and predict gene knockout phenotypes and growth rates quantitatively.

**Acknowledgments** I acknowledge funding through an International Predoctoral Fellowship from the Howard Hughes Medical Institute; I thank Dr. Nathan Price for valuable guidance and James Eddy for help with making some of the figures.

## References

1. Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, Schubert LA, Birditt B, Shay T, Goren A, Zhang X, Smith Z, Deering R, McDonald RC, Cabili M, Bernstein BE, Rinn JL, Meissner A, Root DE, Hacohen N, Regev A (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* 326:257–263
2. Babu MM, Lang B, Aravind L (2009) Methods to reconstruct and compare transcriptional regulatory networks. *Methods Mol Biol* 541:163–180
3. Balazsi G, Heath AP, Shi L, Gennaro ML (2008) The temporal response of the *Mycobacterium tuberculosis* gene regulatory network during growth arrest. *Mol Syst Biol* 4:225
4. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Mol Syst Biol* 3:78
5. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human B cells. *Nat Genet* 37:382–390
6. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2:727–738
7. Becker SA, Palsson BO (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol* 4:e1000082
8. Bonneau R (2008) Learning biological networks: from modules to dynamics. *Nat Chem Biol* 4:658–664
9. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, Longabaugh W, Vuthoori M, Whitehead K, Madar A, Suzuki L, Mori T, Chang DE, Diruggiero J, Johnson CH, Hood L, Baliga NS (2007) A predictive model for transcriptional control of physiology in a free living cell. *Cell* 131:1354–1365
10. Boshoff HI, Myers TG, Copp BR, McNeil MR, Wilson MA, Barry CE (2004) The transcriptional responses of *Mycobacterium tuberculosis* to inhibitors of metabolism: novel insights into drug mechanisms of action. *J Biol Chem* 279:40174–40184
11. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, Holloway E, Kapushesky M, Kemmeren P, Lara GG, Oezcimen A, Rocca-Serra P, Sansone SA (2003) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31:68–71
12. Camacho DM, Collins JJ (2009) Systems biology strikes gold. *Cell* 137:24–26
13. Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, Lasorella A, Aldape K, Califano A, Iavarone A (2010) The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463:318–325
14. Chandrasekaran S, Ament SA, Eddy JA, Rodriguez-Zas SL, Schatz BR, Price ND, Robinson GE (2011) Behavior-specific changes in transcriptional modules lead to distinct and predictable neurogenomic states. *Proc Natl Acad Sci U S A* 108:18020–18025
15. Chandrasekaran S, Price ND (2010) Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* 107:17845–17850
16. Colijn C, Brandes A, Zucker J, Lun DS, Weiner B, Farhat MR, Cheng TY, Moody DB, Murray M, Galagan JE (2009) Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS Comput Biol* 5:e1000489
17. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429:92–96
18. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. *J Theor Biol* 213:73–88
19. Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics* 24:2044–2050

20. Davidson EH, Rast JP, Oliveri P, Ransick A, Caletani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Rust AG, Pan Z, Schilstra MJ, Clarke PJ, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H (2002) A genomic regulatory network for development. *Science* 295:1669–1678
21. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A* 104:1777–1782
22. Edelman LB, Chandrasekaran S, Price ND (2010) Systems biology of embryogenesis. *Reprod Fertil Dev* 22:98–105
23. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210
24. Efron B (2002) Least angle regression. Department of Biostatistics, Stanford University, Stanford, CA
25. Faith JJ, Driscoll ME, Fusaro VA, Cosgrove EJ, Hayete B, Juhn FS, Schneider SJ, Gardner TS (2008) Many microbe microarrays database: uniformly normalized Affymetrix compendia with structured experimental metadata. *Nucleic Acids Res* 36:D866–D870
26. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS (2007) Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 5:e8
27. Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7:129–143
28. Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotechnol* 26:659–667
29. Friedman N, Linial M, Nachman I, Pe'Er D (2000) Using Bayesian networks to analyze expression data. *J Comput Biol* 7:601–620
30. Goelzer A, Briki FB, Martin-Verstraete I, Noirot P, Bessières P, Aymerich S, Fromion V (2008) Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol* 2:20
31. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431:99–104
32. Henry CS, Dejongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28:977–982
33. Herrgard MJ, Lee BS, Portnoy V, Palsson BO (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* 16:627–635
34. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Ijzerman N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasic I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttila M, Klipp E, Palsson BO, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* 26:1155–1160
35. Hood L, Perlmutter R (2004) The impact of systems approaches on biological problems in drug discovery. *Nat Biotechnol* 22:1215–1217
36. Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5:e12776
37. Jensen PA, Papin JA (2011) Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* 27:541–547
38. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:D480–D484

39. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B Jr, Assad-Garcia N, Glass JI, Covert MW (2012) A whole-cell computational model predicts phenotype from genotype. *Cell* 150:389–401
40. Kauffman KJ, Prakash P, Edwards JS (2003) Advances in flux balance analysis. *Current Opin Biotechnol* 14:491–496
41. Kitano H (2002) Systems biology: a brief overview. *Science* 295:1662–1664
42. Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, Geschwind DH (2009) Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462:213–217
43. Lee JM, Gianchandani EP, Eddy JA, Papin JA (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol* 4:e1000086
44. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799–804
45. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10:291–305
46. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng* 5:264–76
47. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, Camacho DM, Allison KR, Aderhold A, Allison KR, Bonneau R, Camacho DM, Chen Y, Collins JJ, Cordero F, Costello JC, Crane M, Dondelinger F, Drton M, Esposito R, Foygel R, de la Fuente A, Gertheiss J, Geurts P, Greenfield A, Grzegorzczak M, Haury AC, Holmes B, Hothorn T, Husmeier D, Huynh-Thu VA., Irrthum, A., Kellis M, Karlebach G, Kuffner R, Lebre S, de Leo V, Madar A, Mani S, Marbach D, Mordelet F, Ostrer H, Ouyang Z, Pandya R, Petri T, Pinna A, Poultney CS, Prill RJ., Reznay S, Ruskin HJ, Saeyes Y, Shamir R, Sirbu A, Song M, Soranzo N, Statnikov A, Stolovitzky G, Vega N, Vera-Licona P, Vert JP, Visconti A, Wang H, Wehenkel L, Windhager L, Zhang Y, Zimmer R, Kellis M, Collins JJ, Stolovitzky G (2012a) Wisdom of crowds for robust gene network inference. *Nat Methods* 9:796–804
48. Marbach D, Prill RJ, Schaffter T, Mattiussi C, Floreano D, Stolovitzky G (2010) Revealing strengths and weaknesses of methods for gene network inference. *Proc Natl Acad Sci USA* 107:6286–6291
49. Marbach D, Roy S, Ay F, Meyer PE, Candeias R, Kahveci T, Bristow CA, Kellis M (2012b) Predictive regulatory models in *Drosophila melanogaster* by integrative inference of transcriptional networks. *Genome Res* 22:1334–1349
50. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Califano A (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinf* 7(Suppl 1):S7
51. Mo ML, Palsson BO, Herrgard MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol* 3:37
52. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. *Nat Neurosci* 11:1271–1282
53. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol Syst Biol* 7:535
54. Price ND, Papin JA, Schilling CH, Palsson BO (2003) Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol* 21:162–169
55. Price ND, Reed JL, Palsson BO (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2:886–897
56. Raman K, Rajagopalan P, Chandra N (2005) Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLoS Comput Biol* 1:e46
57. Robinson GE (2004) Genomics. Beyond nature and nurture. *Science* 304:397–399
58. Robinson GE, Fernald RD, Clayton DF (2008) Genes and social behavior. *Science* 322:896–900



59. Rodionov DA (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* 107:3467–3497
60. Roy S, Ernst J, Kharchenko PV, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow CA, Ma L, Lin MF, Washietl S, Arshinoff BI, Ay F, Meyer PE, Robine N, Washington NL, di Stefano L, Berezikov E, Brown CD, Candeias R, Carlson JW, Carr A, Jungreis I, Marbach D, Sealfon R, Tolstorukov MY, Will S, Alekseyenko AA, Artieri C, Booth BW, Brooks AN, Dai Q, Davis CA, Duff MO, Feng X, Gorchakov AA, Gu T, Henikoff JG, Kapranov P, Li R, Macalpine HK, Malone J, Minoda A, Nordman J, Okamura K, Perry M, Powell SK, Riddle NC, Sakai A, Samsonova A, Sandler JE, Schwartz YB, Sher N, Spokony R, Sturgill D, van Baren M, Wan KH, Yang L, Yu C, Feingold E, Good P, Guyer M, Lowdon R, Ahmad K, Andrews J, Berger B, Brenner SE, Brent MR, Cherbas L, Elgin SC, Gingeras TR, Grossman R, Hoskins RA, Kaufman TC, Kent W, Kuroda MI, Orr-Weaver T, Perimon N, Pirrotta V, Posakony JW, Ren B, Russell S, Cherbas P, Graveley BR, Lewis S, Micklem G, Oliver B, Park PJ, Celniker SE, Henikoff S, Karpen GH, Lai EC, Macalpine DM, Stein LD, White KP, Kellis M (2010) Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* 330:1787–1797
61. Salgado H, Gama-Castro S, Martinez-Antonio A, Diaz-Peredo E, Sanchez-Solano F, Peralta-Gil M, Garcia-Alonso D, Jimenez-Jacinto V, Santos-Zavaleta A, Bonavides-Martinez C, Collado-Vides J (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res* 32:D303–D306
62. Sassetti CM, Boyd DH, Rubin EJ (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 48:77–84
63. Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 26:1003–1010
64. Shlomi T, Eisenberg Y, Sharan R, Ruppin E (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol* 3:101
65. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, Laxman B, Mehra R, Lonigro RJ, Li Y, Nyati MK, Ahsan A, Kalyana-Sundaram S, Han B, Cao X, Byun J, Omenn GS, Ghosh D, Pennathur S, Alexander DC, Berger A, Shuster JR, Wei JT, Varambally S, Beecher C, Chinnaiyan AM (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 457:910–914
66. Teixeira MC, Monteiro P, Jain P, Tenreiro S, Fernandes AR, Mira NP, Alenquer M, Freitas AT, Oliveira AL, Sa-Correia I (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Res* 34:D446–D451
67. Winston ML (1987) *The biology of the honey bee*. Harvard University Press, Cambridge



# Index

## A

Activator-inhibitor, 6, 286  
Acyclic, 186, 188, 199, 200  
Acyclicity, 186  
Adaptation, 48, 58, 59  
Adaptive control, 104  
Affine toric variety, 45  
Aikake's information criterion, 160  
Aleatory variability, 7, 287  
Algebraic, 4, 5, 9, 45, 265  
Amino acid, 2, 11, 23, 25, 26, 28, 122, 131, 144–146, 148  
Analyzing subsets of transcriptional regulators influencing expression (ASTRIX), 307, 312–316, 319, 321  
Aristotelian, 7, 287  
Associativity, 221  
Atomic event-systems, 6, 9, 10, 41, 42, 44, 45  
Auto-regulation, 16, 296  
Autocatalytic reaction, 57

## B

Bacillus subtilis (*B. subtilis*), 2, 3, 8, 25, 122, 123, 128, 145  
Basis, 211, 212, 222  
Bayes' theorem, 8, 14, 16, 20–23, 288, 294, 296, 300–303  
Bayesian, 153, 155, 157, 164, 169, 314  
Bayesian estimation, 9, 289  
Betweenness centrality, 213  
Binary operator, 259, 260  
Bind, 189, 190  
Binding, 182  
Binomials, 4, 7, 11

## Biochemical Abstract

Machine (BIOCHAM), 264, 265  
Biological network, 48, 67  
Biomolecular, 228  
Biosynthesis, 228–230, 235  
Bode analysis, 249  
Bode plot, 249, 250  
Boolean, 259, 262  
Boolean Decision Diagram (BDD), 268, 269, 277  
Boolean operator, 260  
Boundary layer, 201, 209  
Bounded, 184–186, 191, 203, 204, 208  
Boundedness, 48, 50, 51, 63–65, 67–69  
Branched, 181, 182, 184, 198

## C

Capacity, 11, 17, 131, 137  
Cascade, 204, 208  
Cauchy's existence theorem, 16  
Causality, 4, 284  
Cell fate, 263–266, 276  
Cell signaling, 85  
Characteristic polynomial, 56, 58, 59, 68, 69  
Chemical reaction network, 26, 48, 66  
Chemical reaction network theory (CRNT), 199  
Chemotherapy, 245  
Cholesterol, 229, 230, 234–236, 243, 252  
Circadian, 229, 247–249, 251  
Closed set, 202  
Cluster, 215–220, 223  
Clustering coefficient, 215, 217  
Co-factor, 1, 15, 121, 135  
Co-metabolite, 1, 15, 121, 135  
Coding, 183, 184, 188–190

Compact set, 196, 200, 202, 203, 209  
 Compartmental system, 89  
 Complementary sigmoid, 52, 57  
 Complex networks, 211, 212  
 Computation, 257–259  
 Computation tree logic (CTL), 260–262, 264, 265, 269, 273, 276  
 Conformation, 189  
 Conjunction, 259  
 Conservation class, 5, 6, 13, 38, 40, 41, 43, 44  
 Conservation law, 12, 13, 42–45  
 Constraint-based modeling, 307, 308  
 Consumption, 186, 191  
 Continuous, 5–7, 13–15, 17, 21, 34–37, 39, 50, 125–127, 133, 134, 137, 184, 186, 189, 198, 203, 205, 206, 208  
 Continuous stochastic logic (CSL), 272, 273  
 Continuous time, 230  
 Continuous time Markov chain (CTMC), 272–274, 277  
 Contraction method, 74  
 Contraction principle, 74  
 Control theory, 47  
 Converging input bounded state (CIBS), 203–205  
 Converging input converging state (CICS), 200, 201, 203, 205  
 Convex, 74, 75, 85, 86, 89, 95–99  
 Correlation coefficient, 214  
 Cromwell's rule, 14, 294  
 Cross-talk, 257, 263, 266  
 Cyclic, 98, 99

**D**

Degradation, 184, 189, 190  
 Degree, 213–223  
 Degree distribution, 213, 217–219, 221  
 Derivative, 208  
 Detection, 153–159, 162, 165, 169–171  
 Deterministic, 4, 10, 19, 284, 290, 299  
 Diagonal, 74, 75, 77, 84–86, 90, 99  
 Differentiable, 196, 208  
 Differential entropy, 12, 292  
 Differential equations, 4, 5, 11, 14, 16, 17  
 Diffusive instability, 89  
 Diffusively-coupled, 89  
 Dilution, 187, 189, 190  
 Directed graph, 186  
 Discrete time, 230  
 Disjunction, 259  
 DNA, 1, 23, 121, 144

DNA-damage, 3, 283  
 Dynamic inversion, 103, 111, 112, 115, 117, 118

**E**

Edge, 186, 221  
 Eigenvalue, 48, 54, 57  
 Electrocardiogram (ECG), 154  
 Elementary mode analysis (EMA), 212  
 Elliptic operator, 79  
 Elowitz-Leibler, 103, 104  
 End product control structure (EPCS), 3, 17, 18, 20, 25, 26, 123, 137, 138, 140, 145, 146  
 Energy, 230, 231  
 Energy cycle, 5, 24, 44  
 Entropy, 12, 13, 15, 17, 18, 292, 293, 295, 297, 298  
 Enzymatic networks, 48  
 Enzyme, 3–7, 9–11, 14–20, 24, 25, 27, 28, 123–127, 129–131, 134–140, 144, 146–148, 182–185, 187, 191, 193, 194, 197, 199, 200, 208  
 Enzyme kinetics, 182, 184, 185, 187, 198, 199  
 Epidermal growth factor (EGF), 60  
 Epilepsy, 153, 155, 161, 169  
 Epistemic indeterminacy, 7, 287  
 Epistemic separability, 7, 287  
 Equation, 181, 192, 200  
 Equilibrium, 49, 55, 57–62, 64, 68, 69, 191–193, 198, 199, 201, 204, 207–209  
 Equilibrium point, 6, 8, 9, 11, 24, 27, 28, 38, 40, 44  
 Erdos-Renyi network, 213  
 Escherichia coli (E. coli), 104, 122, 145, 146  
 Event graph, 24, 43  
 Event systems, 4–6, 8–14, 23, 26, 27, 38, 41, 44, 45  
 Expectation maximization (EM), 240  
 Export rate, 187, 188  
 Extended Lyapunov function, 34, 36  
 External coincidence, 248, 251

**F**

Feedback, 59, 61, 62, 104, 105, 108, 110, 111, 116, 117  
 Fitness, 233, 237–239, 249, 251  
 Flow-invariant affine subspaces, 5  
 Fluorescence, 105  
 Fluorescent, 52

Flux, 5, 9–12, 15–20, 23–25, 28, 125, 129–132, 135–140, 144, 145, 148, 149, 193

Flux balance analysis (FBA), 212, 223, 309–311, 317–320

Formula, 259–262

## G

Gaussian, 6, 12, 15, 286, 292, 295

Gene, 182, 183, 188, 190, 198

Gene Regulatory Network (GRN), 270–272

Generalized linear model (GLM), 160, 165, 166

Genetic, 2, 3, 9, 11, 15, 18, 29, 122, 123, 129, 131, 135, 138, 149

Genetic control, 2, 122

Genetic network, 181, 182, 194, 198

Genetic regulation, 181–183, 195

Genome, 52

Genotype, 307, 308, 315, 321

Geodesic, 214, 216, 217, 219–221

Gini coefficient, 214

Global regulation, 3, 18, 26, 123, 138, 146

Globally asymptotic stable (GAS), 193, 198, 200, 201, 203, 209

Goldbeter-Koshland function, 5, 285

Graph, 9, 24, 26, 41, 43, 212–215, 218

Graph analysis, 266, 276

Graph representation, 266, 276

Graphlet, 217, 220

Green's identity, 80

Gutenberg-richter law, 213

## H

Heatmap, 217, 218

Hidden Markov Model (HMM), 153, 155, 156, 171, 172

Hidden state, 235, 238, 239

Hill coefficient, 104, 105

Hill function, 49, 50, 52, 53, 67, 189

Hormone, 228–230, 235, 237

Hull, 98

Hypothesis, 2, 3, 8–10, 12, 15, 18, 19, 22, 282, 283, 288–290, 292, 295, 298, 299, 302

## I

IKK, 245, 246

Import rate, 186, 188

Initial product control structure (IPCS), 3, 17, 18, 24–26, 123, 137, 138, 144–146

Input, 228, 232–235, 240, 243–245, 247–251

Internal coincidence, 248

Intracortical electroencephalogram (iEEG), 154, 155, 159, 161, 164, 170, 171

Invariant, 185, 196, 200, 202

Invariant set, 22, 23

Irreversible, 183, 184, 200

Irreversible enzyme, 1, 11, 12, 14–16, 18, 24, 121, 131, 132, 134–136, 138, 144

Isoenzyme, 3, 15, 16, 123, 135, 136

## J

Jacobian, 48, 55, 57, 68, 69

## K

K-core, 216

Kalman filtering, 20, 300

Knockout, 212, 223

Knockout strain, 212, 223

Krackhardt kite graph, 213

Kronecker product, 89

## L

Lac operon, 66

$\mathcal{L}_1$  adaptive control, 104

Laplacian, 74, 84, 90

Law of mass action, 4, 6, 11

Linear, 189, 205

Linear matrix inequality (LMI), 74, 94, 98

Linear programming, 310

Linear system identification, 227, 235, 244

Linear systems, 227, 232, 244, 247

Linear temporal logic (LTL), 259

Linear time invariant (LTI), 234, 239, 247–251

Linearization, 53, 59, 69

Lipschitz, 184, 186, 189, 205, 208, 209

Local minima, 240, 244, 251

Local regulation, 3, 123

Loop shift transformation, 110

Lure system, 108

Lyapunov function, 5, 26–29, 34, 36, 38

Lyapunov-like function, 64, 67

Lysine, 26–29, 146–149

## M

M. tuberculosis, 320, 321

- Macromolecule, 5, 6, 285, 286  
 Magnetic resonance imaging (MRI), 161  
 Map, 76, 90  
 Markov, 153, 156, 169  
 Markov chain Monte Carlo (MCMC), 20, 300  
 Mass action kinetics, 4–6  
 Matrix measure, 74  
 Mean, 217, 219, 220, 222  
 Mean value theorem, 83, 93  
 Memoryless, 108  
 Metabolic, 1, 2, 10, 12, 14–16, 18, 19, 26, 29, 121, 122, 130, 132, 134–136, 138, 139, 146, 149, 150, 307, 308, 310, 316–321  
 Metabolic network, 1–3, 29, 121–123, 149, 181, 182, 186, 190, 198, 308–311, 316–321  
 Metabolite, 3–7, 10–12, 15, 17, 18, 23–25, 28, 123–127, 130–132, 135, 137, 138, 143–145, 148, 181–184, 189, 192, 194, 198–200  
 Metric, 212–218, 220–223  
 Metzler matrix, 69  
 Michaelis–Menten, 182, 187, 188  
 Microbe, 220–222  
 Minimization of metabolic adjustment (MOMA), 212  
 Mitogen activated protein kinase (MAPK), 104  
 Model checking, 257, 258, 261, 262  
 Model selection, 235, 238, 242, 244, 251  
 Model validation, 235, 249, 251  
 Modular, 29, 149  
 Modularity, 220, 221  
 Molecular, 227, 228  
 Monic monomial, 7  
 Monotone, 69, 108, 109, 111  
 Monotonic, 27  
 Monotonicity, 192, 206, 207  
 Motility, 220, 221  
 mRNA, 103–105, 111, 117, 188  
 Multimodal, 21, 301  
 Multistationarity, 47, 48, 69  
 Mutant, 223  
 Mutation, 263
- N**
- Natural event systems, 5, 6, 9, 13, 23, 24, 26–29, 34, 36, 38, 40, 44  
 Nerve growth factor (NGF), 60  
 Network, 154, 155, 181, 183, 185–188, 192–195, 198, 200, 203, 205, 307–321  
 Network inference, 307, 311  
 Neumann boundary condition, 75, 79–81  
 Neumann eigenvalue, 74–76  
 Nf- $\kappa$ B, 244, 246  
 Nf- $\kappa$ B, 229, 244–246  
 Node, 103, 105, 110, 111, 117, 213–215, 217  
 Nonlinearity, 227, 243, 251, 252
- O**
- Oncogenesis, 245  
 Operand, 259, 262  
 Operator, 259, 260  
 Optimal control, 153, 155, 169  
 Optimal detection policy (ODP), 155, 159, 162, 167, 170, 171  
 Ordinary differential equation (ODE), 181, 182, 185  
 Orthogonal, 76, 81, 90–92  
 Orthogonal projection, 76, 90  
 Oscillation, 47, 48, 56, 57, 103, 104, 109, 110, 113–117  
 Oscillator, 104, 114  
 Over-fitting, 235, 239, 244, 251  
 Over-parameterisation, 238
- P**
- Parameter, 228, 230–234, 237, 238, 240, 243, 245, 248  
 Pareto law, 213  
 Pareto-efficient, 19, 299  
 Partial differential equation (PDE), 74–77, 90, 99  
 Path, 258, 260–262  
 Pathway, 2, 3, 5, 7–9, 11, 12, 14–19, 23, 26, 27, 29, 48, 60, 62, 67, 122, 123, 125, 127–129, 131, 132, 134–139, 144, 146, 147, 149, 150, 183, 184, 188, 190, 198, 199  
 Pattern formation, 74  
 Permutation, 99  
 Persistent network, 62  
 Perturbation, 181  
 Phase transition, 213  
 Phenotype, 307, 308, 310, 315, 316, 319–321  
 Phosphorylation, 4, 284  
 PID control, 114, 116  
 Poisson degree distribution, 213  
 Positive, 184, 191, 193  
 Positive definite, 185, 191  
 Positivity, 184  
 Power law, 213

Prediction error method (PEM), 232–234, 237, 239, 240  
 Pregnenolone, 235, 237  
 Principal component analysis (PCA), 217–220  
 Principle of indifference, 15, 295  
 Principle of maximum entropy, 15, 16, 18, 295, 296, 298  
 Probabilistic regulation of metabolism (PROM), 307, 311, 316, 318–321  
 Probability, 153, 155–158, 160, 164, 167–169, 171  
 Production, 182, 183, 186  
 Progesterone, 229–231, 233, 235, 236, 243, 252  
 Progesterone synthesis, 229  
 Projection, 76, 90  
 Promoter, 104, 105, 111, 189, 190, 198  
 Proportional+integral (PI), 111, 115  
 Proposition, 258  
 Protein, 49, 60, 61, 65, 67, 103–105, 111, 117  
 Purine, 23, 24, 144

## Q

Quantitative network (QN), 270, 271  
 Quasi-convex, 96

## R

Reaction rate, 184, 185  
 Reaction-diffusion, 74, 75, 78, 84, 86, 99  
 Realization, 54, 57, 65  
 Reduced system, 194, 196, 198, 201, 202  
 Regulatory, 307, 308, 310–313, 315–321  
 Regulatory network, 1, 2, 26, 29, 121, 122, 146, 149  
 Relative entropy, 13, 14, 17, 293, 294, 297  
 Reporter, 52  
 Repressilator, 103–106, 109, 110, 112–117  
 Reversible, 200  
 Reversible enzymes, 5, 125  
 Ribosome, 189  
 Riboswitch, 27, 147  
 Rich-club coefficient, 215  
 Right non-negative (RNN), 18, 19  
 Ring oscillator, 97, 98  
 RNA, 6, 23, 27, 28, 49, 52, 126, 144, 147, 148  
 RNA polymerase, 6, 28, 126, 148  
 Robustness, 48, 53, 54, 69  
 Root, 194, 196, 202, 209  
 Ruth–Hurwitz, 58

## S

S-metric, 213  
 Satisfiability modulo theory (SMT), 269, 277  
 Saturation, 50, 61, 62  
 SBML, 275  
 Scale free, 213, 216  
 Second law of thermodynamics, 5, 27, 44  
 Seizure, 153–155, 159, 161, 162, 164, 165, 167, 168  
 Self-information, 10, 11, 13, 290, 291, 293  
 Sequence diagram, 264  
 Sequential Monte Carlo method, 21, 301  
 Set invariance, 53  
 Shannon-Khinchin axiom, 12, 292  
 Sigmoid, 51, 52  
 Singular value decomposition, 159, 163  
 Singularly perturbed system, 181  
 Smooth, 184, 186, 189, 201, 209  
 Spatial uniformity, 74, 75, 89, 94, 99  
 Spatio-temporal dynamics, 74  
 Stability, 103, 105, 106, 108–110, 113, 117, 201, 205  
 Standard deviation, 217, 219, 220  
 State, 228, 233–235, 238, 239, 243, 245, 248, 252  
 State space, 233  
 Steady-state, 1, 2, 7–9, 11–17, 121, 122, 127–129, 131–137, 212  
 Steroid, 229, 235  
 Stochastic differential equation (SDE), 8, 288  
 Stochastic process, 239  
 Stoichiometric coefficient, 5, 44, 186  
 Stoichiometric constraint, 308, 310  
 Stoichiometry, 186  
 Strictly increasing, 185, 191, 192, 208  
 String theory, 4  
 Structural analysis, 47, 48, 51  
 Structural property, 48, 54, 57  
 Subspace, 76, 90  
 Substrate, 182  
 Subsystem, 200, 205–208  
 Surjective, 95, 96  
 Symmetric, 75, 78–80, 82, 84–93, 95, 96  
 Synchronization, 160, 164  
 Synchrony, 99  
 Synthesis, 184, 188–190  
 Synthetic accessibility, 212, 223  
 System, 227–235, 237–239, 243–248, 251  
 System identification, 227–229, 235, 239, 243–245, 248, 251, 252

**T**

Taylor series, 14, 15, 18–20  
Temporal logic, 258, 259  
Thermodynamic constraint, 318, 319  
Tikhonov's theorem, 181, 194, 200–202, 209  
Time delay, 267  
Time scale separation, 181–183, 190, 193–195, 198, 199, 201  
Topological analysis, 212, 223  
Topology, 48, 60, 212, 220  
Tracking controller, 103, 110, 111, 116, 117  
Transcription, 8, 16, 288, 296  
Transcription factor (TF), 189, 190  
Transcriptional network, 104  
Transcriptional regulatory network (TRN), 310, 312–317, 320  
Transcriptional repression, 49, 53, 54  
Transfer function, 109  
Transition, 258–262  
Transition system, 258, 259, 261  
Trif, 246, 247

Tryptophan biosynthesis, 18, 298  
Tunable, 104, 109, 110, 113, 116, 117  
Turbulence, 213

**U**

Unary operator, 259, 260  
Unimodal, 21, 301  
Unprime, 268  
Uptake rate, 212

**V**

Vertex, 186, 215

**Z**

Zames-Falb multiplier, 103, 105, 108, 111, 117  
Zero deficiency theorem, 68  
Zipf law, 213